Relaxing Integrity Requirements for Attack-Resilient Cyber-Physical Systems

Ilija Jovanov, and Miroslav Pajic, *Member, IEEE*

Abstract—**The increase in network connectivity has also resulted in several high-profile attacks on cyber-physical systems. An attacker that manages to access a local network could *remotely* affect control performance by tampering with sensor measurements delivered to the controller. Recent results have shown that with network-based attacks, such as *Man-in-the-Middle* attacks, the attacker can introduce an unbounded state estimation error if measurements from a suitable subset of sensors contain false data when delivered to the controller. While these attacks can be addressed with the standard cryptographic tools that ensure data integrity, their continuous use would introduce significant communication and computation overhead. Consequently, we study effects of intermittent data integrity guarantees on system performance under stealthy attacks. We consider linear estimators equipped with a general type of residual-based intrusion detectors (including e.g., widely used χ^2 detectors), and show that even when integrity of sensor measurements is enforced only intermittently, the attack impact is significantly limited; specifically, the state estimation error is bounded or the attacker cannot remain stealthy. Furthermore, we present methods to: (1) evaluate the effects of any given integrity enforcement policy in terms of reachable state-estimation errors for any type of stealthy attacks, and (2) design an enforcement policy that provides the desired estimation error guarantees under attack. Finally, on three automotive case studies we show that even with less than 10% of authenticated messages we can ensure satisfiable control performance in the presence of attacks.**

Index Terms—**Attack-resilient state estimation, attack detection, Kalman filtering, cyber-physical systems security, linear systems.**

I. INTRODUCTION

S EVERAL high-profile incidents have recently exposed vulnerabilities of cyber-physical systems (CPS) and drawn attention to the challenges of providing security guarantees as part of their design. These incidents cover a wide range of application domains and system complexity, from attacks on large-scale infrastructure such as the 2016 breach of Ukrainian power-grid [42], to the StuxNet virus attack on an industrial

I Jovanov and M. Pajic are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708 USA. E-mail: {ilija.jovanov, miroslav.pajic}@duke.edu.

SCADA system [13], as well as attacks on controllers in modern cars (e.g., [2]) and unmanned arial vehicles [31].

There are several reasons for such number of security related incidents affecting control of CPS. The tight interaction between information technology and physical world has greatly increased the attack vector space. For instance, an adversarial signal can be injected into measurements obtained from a sensor, using non-invasive attacks that modify the sensor's physical environment; as shown in attacks on GPS-based navigation systems [9], [41]. Even more important reason is network connectivity that is prevalent in CPS. An attacker that manages to access a local control network could *remotely* affect control performance by tampering with sensor measurements and actuator commands in order to force the plant into any desired state, as illustrated in [34]. From the controls perspective, attacks over an internal system network, such as the *Man-in-the-Middle* (MitM) attacks where the attacker inserts messages anywhere in the sensors→controllers→actuators pathway, can be modeled as additional malicious signals injected into the control loop via the system's sensors and actuators [38].

While the interaction with the physical world introduces new attack surfaces, it also provides opportunities to improve system resilience agains attacks. The use of control techniques that employ a physical model of the system's dynamics for attack detection and attack-resilient state estimation has drawn significant attention in recent years (e.g., [3], [5], [24]–[28], [32], [35], [37]–[40], and a recent survey [17]). One line of work is based on the use of unknown input observers (e.g., [28], [37]) and non-convex optimization for resilient estimation (e.g., [3], [25]), while another focuses on attack-detection and estimation guarantees in systems with standard Kalman filter-based state estimators (e.g., [10]–[12], [21]–[24]). In the later, estimation residue-based failure detectors, such as χ^2 [22], [23], cumulative sum (CUSUM) [30], and sequential probability ratio test (SPRT) detectors [12], are employed for intrusion detection. Still, irrelevant of the used attack detection mechanism, after compromising a suitable subset of sensors, an intelligent attacker can significantly degrade control performance while remaining undetected (i.e., stealthy). For example, for resilient state estimation techniques (e.g., [3], [25]), measurements from at least half of the sensors should not be tampered with [3], [33], while [11], [22] capture attack requirements for Kalman filter-based estimators. Such conservative results are caused by the common initial assumption that once a sensor or its communication to the estimator is compromised, all values received from the sensor can be potentially corrupted – i.e., integrity of the data received
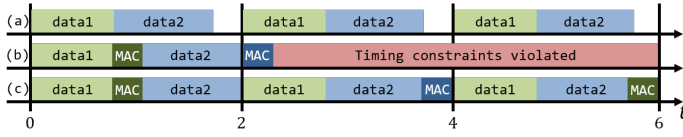
Fig. 1. Communication schedule for periodic messages (with period $T_s = 2$) from two sensors over a shared network: (a) a feasible schedule for non-authenticated messages (i.e., when MAC bits are not attached to transmitted packets); (b) there is no feasible schedule when all messages are authenticated; and (c) if data integrity is only intermittently enforced (e.g., by adding MAC bits only to every other packet), scheduling of the messages becomes feasible.

from these sensors cannot be guaranteed.

On the other hand, most of network-based attacks, including MitM attacks, can be avoided with the use of standard cryptographic tools. For example, to authenticate data and ensure integrity of received communication packets, a common approach is to add a message authentication code (MAC) to the transmitted sensor measurements. Therefore, data integrity requirements can be imposed by the continuous use of MACs in all transmissions from a sufficient subset of sensors. However, the overhead caused by the continuous computation and communication of authentication codes can limit their use. For instance, adding MAC bits to networked control systems that employ Controller Area Networks (CAN) may not be feasible due to the message length limitation (e.g., only 64 payload bits per packet in the basic CAN protocol), while splitting them into several communication packets significantly increases the message transmission time [16]. To illustrate this, consider two sensors periodically transmitting measurements over a shared network. As presented in Figure 1(a), without authentication (i.e., if transmitted data contain no MAC bits) the communication packets will be schedulable but the system would be vulnerable to false-data injection attacks. Yet, if all measurements from both sensors are authenticated, with the increase in the packet size due to authentication overhead, it is not possible to schedule transmissions from both sensors in every communication frame (Figure 1(b)). Finally, a feasible schedule exists if MAC bits are attached to every other measurement packet transmitted by each sensor (Figure 1(c)).

Consequently, in this paper we focus on state estimation in systems with *intermittent data integrity guarantees* for sensor measurements delivered to the estimator. Specifically, we study the performance of linear filters equipped with residual-based intrusion detectors in the presence of attacks on sensor measurements. We build on the system model from [11], [22], [23] by capturing that the use of authentication mechanisms in intermittent time-points ensures that sensor measurements received at these points are valid. To keep our discussion and results general, we consider a wide class of detection functions that encompasses commonly used detectors, such as χ^2 and log-likelihood ratio detectors. We show that even when integrity of communicated sensor data is enforced only intermittently and the attacker is fully aware of the times of the enforcement, the attack impact gets significantly limited; concretely, either the state estimation error remains bounded or the attacker cannot remain stealthy. This holds even when communication from *all* sensors to the estimator can be compromised as well as in any other case where otherwise (i.e., without integrity enforcements) an unbounded estimation error can be introduced.

Furthermore, to facilitate the use of intermittent data integrity enforcement for control of CPS in the presence of network-based attacks, we introduce an analysis and design framework that addresses two challenges. First, we introduce techniques to evaluate the effects of any given integrity enforcement policy in terms of reachable state-estimation errors for any type of stealthy attacks. Note that methods to evaluate potential state estimation errors due to attacks are considered in [12], [22], [23]. However, given that the previous work considers system architectures without intermittent use of authentication, these techniques result in overly conservative estimates of reachable regions or they cannot capture the effects of intermittent integrity guarantees on the estimation error. Second, we present a method to design an enforcement policy that provides the desired estimation error guarantees for any attack signal inserted via compromised sensors. The developed framework also facilitates tradeoff analysis between the allowed estimation error and the rate at which data integrity should be enforced – i.e., the required system resources such as communication bandwidth as we have presented in [14].

The rest of the paper is organized as follows. In Section II, we introduce the problem, including the system and attack models. In Section III, we analyze the impact of stealthy attacks in systems without integrity enforcements and formally define intermittent integrity enforcement policies. Section IV focuses on state estimation guarantees when data integrity is at least intermittently enforced. We then introduce a methodology to analyze effects of integrity enforcement policies as well as design suitable policies that ensure the desired estimation error even in the presence of attacks (Section V). Finally, in Section VI, we present case studies that illustrate effectiveness of our approach, before providing final remarks in Section VII.

A. Notation and Terminology

The transpose of matrix $\mathbf{A}$ is specified as $\mathbf{A}^T$, while the ith element of a vector $\mathbf{x}_k$ is denoted by $\mathbf{x}_{k,i}$.[1] Moore-Penrose pseudoinverse of matrix $\mathbf{A}$ is denoted as $\mathbf{A}^\dagger$. In addition, $\|\mathbf{A}\|_i$ denotes the i-norm of a matrix $\mathbf{A}$ and, for a positive definite matrix $\mathbf{Q}$, $\|\Delta\mathbf{z}_k\|_{\mathbf{Q}^{-1}} = \|\mathbf{Q}^{-1/2}\Delta\mathbf{z}_k\|_2$. $null(\mathbf{A})$ denotes the null space of the matrix. Also, $diag(\cdot)$ indicates a square matrix with the quantities inside the brackets on the diagonal, and zeros elsewhere, while $BlckDiag(\cdot)$ denotes a block-diagonal operator. We denote positive definite and positive semidefinite matrix $\mathbf{A}$ as $\mathbf{A} \succ 0$ and $\mathbf{A} \succeq 0$, respectively, while $det(\mathbf{A})$ stands for the determinant of the matrix. Also, $\mathbf{I}_p$ denotes the p-dimensional identity matrix, and $\mathbf{0}_{p \times q}$ denotes $p \times q$ matrix of zeroes. We use $\mathbb{R}$, $\mathbb{N}$ and $\mathbb{N}_0$ to denote the sets of reals, natural numbers and nonnegative integers, respectively. As most of our analysis considers bounded-input systems, we refer to any eigenvalue λ as *unstable eigenvalue* if $|\lambda| \geq 1$.

For a set $\mathcal{S}$, we use $|\mathcal{S}|$ to denote the cardinality (i.e., size) of the set. In addition, for a set $\mathcal{K} \subset \mathcal{S}$, with $\mathcal{K}^{\complement}$ we denote the complement set of $\mathcal{K}$ with respect to $\mathcal{S}$ – i.e., $\mathcal{K}^{\complement} = \mathcal{S} \setminus \mathcal{K}$.

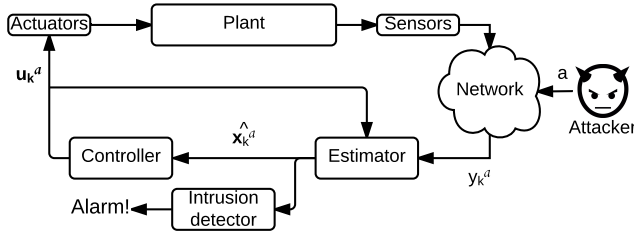[1] We use bold letter to denote matrices and vectors (i.e., non-scalars).

Fig. 2. System architecture – by launching *Man-in-the-Middle* (MitM) attacks, the attacker can inject adversarial signals into plant measurements obtained from system sensors.

Projection vector $\mathbf{i}_j^T$ denotes the row vector (of the appropriate size) with a 1 in its jth position being the only nonzero element of the vector. For a vector $\mathbf{y} \in \mathbb{R}^p$, we use $\mathbf{P}_{\mathcal{K}}\mathbf{y}$ to denote the projection from the set $\mathcal{S} = \{1, ..., p\}$ to set $\mathcal{K}$ ($\mathcal{K} \subseteq \mathcal{S}$) by keeping only elements of $\mathbf{y}$ with indices from $\mathcal{K}$.[2] Finally, *the support* of the vector $\mathbf{v} \in \mathbb{R}^p$ is the set $\text{supp}(\mathbf{v}) = \{i \mid \mathbf{v}_i \neq 0\} \subseteq \{1, 2, ..., p\}$.

II. PROBLEM DESCRIPTION

Before introducing the problem formulation, we describe the considered system and its architecture (shown in Figure 2), as well as the attacker model.

A. System Model without Attacks

We consider an observable linear-time invariant (LTI) system whose evolution without attacks can be represented as

$$
\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \\
\mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{v}_k
\end{aligned}
\tag{1}
$$

where $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{u}_k \in \mathbb{R}^m$ denote the plant's state and input vectors, at time k, while the plant's output vector $\mathbf{y}_k \in \mathbb{R}^p$ contains measurements provided by p sensors from the set $\mathcal{S} = \{s_1, s_2, ..., s_p\}$. Accordingly, the matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ have suitable dimensions. Also, $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^p$ denote the process and measurement noise; we assume that $\mathbf{x}_0$, $\mathbf{w}_k$, and $\mathbf{v}_k$ are independent Gaussian random variables.

Furthermore, the system is equipped with an estimator in the form of a Kalman filter. Given that the Kalman gain usually converges in only a few steps, to simplify the notation we assume that the system is in steady state before the attack. Hence, the Kalman filter estimate $\hat{\mathbf{x}}_k$ is updated as

$$
\hat{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k + \mathbf{K}(\mathbf{y}_{k+1} - \mathbf{C}(\mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k)) \tag{2}
$$

$$
\mathbf{K} = \mathbf{\Sigma}\mathbf{C}^T(\mathbf{C}\mathbf{\Sigma}\mathbf{C}^T + \mathbf{R})^{-1}, \tag{3}
$$

where $\mathbf{\Sigma}$ is the estimation error covariance matrix, and $\mathbf{R}$ is the sensor noise covariance matrix. Also, the residue $\mathbf{z}_k \in \mathbb{R}^p$ at time k and its covariance matrix $\mathbf{Q}$ are defined as

$$
\begin{aligned}
\mathbf{z}_k &= \mathbf{y}_k - \mathbf{C}(\mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_{k-1}), \\
\mathbf{Q} &= E\{\mathbf{z}_k\mathbf{z}_k^T\} = \mathbf{C}\mathbf{\Sigma}\mathbf{C}^T + \mathbf{R}.
\end{aligned}
\tag{4}
$$

Finally, the state estimation error $\mathbf{e}_k$ is defined as the difference between the plant's state $\mathbf{x}_k$ and Kalman filter estimate $\hat{\mathbf{x}}_k$ as

$$
\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k. \tag{5}
$$

In addition to the estimator, we assume that the system is equipped with an intrusion detector. We consider a general case where the detection function g_k of the intrusion detector is defined as

$$
g_k = \sum_{i=k-\mathcal{T}+1}^{k} c_{(i-k+\mathcal{T})}\mathbf{z}_i^T\mathbf{Q}^{-1}\mathbf{z}_i. \tag{6}
$$

Here, $\mathcal{T}$ is the length of the detector's time window, and c_i for $i = 1, ..., \mathcal{T}$ are predefined non-negative coefficients, with $c_{\mathcal{T}}$ being strictly positive. The above formulation captures both fixed window size detectors, where $\mathcal{T}$ is a constant, as well as detectors where the time window size $\mathcal{T}$ satisfies $\mathcal{T} = k$. Also, the definition of the detection function g_k covers a wide variety of commonly used intrusion detectors, such as χ^2 and other detectors previously considered in these scenarios [10]–[12], [18]–[20], [23], [30]. The alarm is triggered when the value of the detection function g_k satisfies

$$
g_k > threshold, \tag{7}
$$

and the probability of the alarm at time k can be captured as

$$
\beta_k = P(g_k > threshold). \tag{8}
$$

B. Attack Model

We assume that the attacker is capable of launching MitM attacks on communication channels between a subset of the plant's sensors $\mathcal{K} \subseteq \mathcal{S}$ and the estimator; for instance, by secretly relaying corresponding altered communication packets. However, we do not assume that the set $\mathcal{K}$ is known to the system or system designers. Thus, to capture the attacker's impact on the system, the system model from (1) becomes

$$
\begin{aligned}
\mathbf{x}_{k+1}^a &= \mathbf{A}\mathbf{x}_k^a + \mathbf{B}\mathbf{u}_k^a + \mathbf{w}_k \\
\mathbf{y}_k^a &= \mathbf{C}\mathbf{x}_k^a + \mathbf{v}_k + \mathbf{a}_k.
\end{aligned}
\tag{9}
$$

Here, $\mathbf{x}_k^a$ and $\mathbf{y}_k^a$ denote the state and plant outputs in the presence of attacks, from the perspective of the estimator, since in the general case they differ from the plant's state and outputs of the non-compromised system. In addition, $\mathbf{a}_k$ denotes the signals injected by the attacker at time k starting from $k = 1$ (i.e., $\mathbf{a}_0 = \mathbf{0}$);[3] to model MitM attacks on communication between the sensors from set $\mathcal{K}$ and the estimator, we assume that $\mathbf{a}_k$ is a sparse vector from $\mathbb{R}^p$ with support in the set $\mathcal{K}$ – i.e., $\mathbf{a}_{k,i} = 0$ for all $i \in \mathcal{K}^C$ and $k > 0$.[4]

We consider the following ***threat model***.

(1) The attacker has full knowledge of the system – in addition to knowing the dynamical model of the plant, employed Kalman filter, and detector, the attacker is aware of all potential

---

[2]Formally, $\mathbf{P}_{\mathcal{K}} = \begin{bmatrix} \mathbf{i}_{k_1} | \cdots | \mathbf{i}_{k_{|\mathcal{K}|}} \end{bmatrix}^T$, where $\mathcal{K} = \{s_{k_1}, ..., s_{k_{|\mathcal{K}|}}\} \subseteq S$ and $k_1 < k_2 < ... < k_{|\mathcal{K}|}$.

[3]More details about why the attacker does not insert attack at step $k = 0$ can be found in Remark 1.

[4]Although a sensor itself may not be directly compromised with MitM attacks, but rather communication between the sensor and estimator, we will also refer to these sensors are *compromised sensors*. In addition, in this work we sometimes abuse the notation by using $\mathcal{K}$ to denote both the set of compromised sensors and the set of indices of the compromised sensors.

security mechanism used in communication. Specifically, we consider systems that use standard methods for message authentication to ensure data integrity, and assume that the attacker is aware at which time points data integrity will be enforced. Thus, to avoid being detected, the attacker will not launch attacks in these steps and will also take into account these integrity enforcements in planning its attacks (as described in Section III).[5] Since we model our system such that attacks start at $k = 1$, this further implies that at $k = 1$ data integrity is not enforced, as otherwise the attacker would not be able to insert false data.

(2) The attacker has the required computation power to calculate suitable attack signals, while planning ahead as needed. (S)he also has the ability to inject *any* signal using communication packets mimicking sensors from the set $\mathcal{K}$, except at times when data integrity is enforced. For instance, when MACs are used to ensure data integrity and authenticity of communication packets, our assumption is that the attacker does not know the shared secret key used to generate the MACs.

The goal of the attacker is to design attack signal $\mathbf{a}_k$ such that it *maximizes the error of state estimation* while ensuring that *the attack remains stealthy*. To formally capture this objective and the stealthiness constraint, we denote the state estimation, residue, and estimation error of the compromised system by $\hat{\mathbf{x}}_k^a$, $\mathbf{z}_k^a$, and $\mathbf{e}_k^a$, respectively. Thus, the attacker's aim is to maximize $\mathbf{e}_k^a$, while ensuring that the increase in the probability of alarm is not significant. We also define as

$$\Delta \mathbf{e}_k = \mathbf{e}_k^a - \mathbf{e}_k, \qquad \Delta \mathbf{z}_k = \mathbf{z}_k^a - \mathbf{z}_k,$$

the change in the estimation error and residue, respectively, caused by the attacks. From (1) and (9), the evolution of these signals can be captured as a dynamical system Ξ of the form

$$\Delta \mathbf{e}_{k+1} = (\mathbf{A} - \mathbf{K}\mathbf{C}\mathbf{A})\Delta \mathbf{e}_k - \mathbf{K}\mathbf{a}_{k+1}, \tag{10}$$

$$\Delta \mathbf{z}_k = \mathbf{C}\mathbf{A}\Delta \mathbf{e}_{k-1} + \mathbf{a}_k, \tag{11}$$

with $\Delta \mathbf{e}_0 = 0$.

Remark 1: From the above equations, the first attack vector to affect the change in estimation error is $\mathbf{a}_1$. Thus, without loss of generality, we assume that the attack starts at $k = 1$ (i.e., $\mathbf{a}_i = 0$, $for\ all\ i \leq 0$). This also implies that $\Delta \mathbf{z}_0 = 0$. Note that the above dynamical system is noiseless (and deterministic), with input $\mathbf{a}_k$ controlled by the attacker. Therefore, since $E[\mathbf{e}_k] = \mathbf{0}$ for the non-compromised system in steady state, it follows that

$$\Delta \mathbf{e}_k = E[\Delta \mathbf{e}_k] = E[\mathbf{e}_k^a]. \tag{12}$$

Given that $\Delta \mathbf{e}_k$ provides expectation of the state estimation error under the attack, this signal can be used to evaluate the impact that the attacker has on the system.[6] Thus, we specify the objective of the attacker as to **maximize the expected state estimation error** (e.g., $\|\Delta \mathbf{e}_k\|_2$). This is additionally justified

---

[5]In Section IV, we will also consider the case where the attacker has limited knowledge of the system's use of security mechanisms.

[6]For this reason, and to simplify our presentation, in the rest of the paper we will sometimes refer to $\Delta \mathbf{e}_k$ as *the (expected) state estimation error* instead of *the change of the state estimation error* caused by attacks.

by the fact that since $\mathbf{a}_k$ is controlled by the attacker (i.e., deterministic to simplify of our presentation), which implies

$$Cov(\mathbf{e}_k^a) = Cov(\mathbf{e}_k) = \mathbf{\Sigma}. \tag{13}$$

To capture the attacker's stealthiness requirements, we use the probability of alarm in the presence of an attack

$$\beta_k^a = P(g_k^a > threshold), \quad \text{where} \tag{14}$$

$$g_k^a = \sum_{i=k-\mathcal{T}+1}^{k} c_{(i-k+\mathcal{T})} \mathbf{z}_i^{aT} \mathbf{Q}^{-1} \mathbf{z}_i^a. \tag{15}$$

Therefore, to ensure that attacks remain stealthy, the attacker's *stealthiness constraint* in each step k is to maintain

$$\beta_k^a \leq \beta_k + \varepsilon, \tag{16}$$

for a small predefined value of $\varepsilon > 0$.

C. Problem Formulation

As we will present in the next section, for a large class of systems, a stealthy attacker can easily introduce an unbounded state estimation error by compromising communication between some of the sensors and the estimator. On the other hand, existing communication protocols commonly incorporate security mechanisms (e.g., MAC) that can ensure integrity of delivered sensor measurements. Specifically, this means that the system could enforce $\mathbf{a}_{k,i} = 0$ for some sensor s_i, or $\mathbf{a}_k = \mathbf{0}$ if integrity for all transmitted sensor measurements is enforced at some time-step k. However, as we previously described, the integrity enforcement comes at additional communication and computation cost, effectively preventing their continuous use in resource constrained CPS.

Consequently, we focus on the problem of evaluating the impact of stealthy attacks in systems with intermittent (i.e., occasional) use of data integrity enforcement mechanisms.[7] Specifically, we will address the following problems:

- Can the attacker introduce unbounded state estimation errors in systems with intermittent integrity guarantees?
- How to efficiently evaluate the impact of intermittent integrity enforcement policies on the induced state estimation errors in the presence of a stealthy attacker?
- How to design a non-overly conservative development framework that incorporates guarantees for estimation degradation under attacks into design of suitable integrity enforcement policies?

III. IMPACT OF STEALTHY ATTACKS ON STATE ESTIMATION ERROR

To capture the impact of stealthy attacks on the system, we start with the following definition.

Definition 1: The set of all stealthy attacks up to time k is

$$\mathcal{A}_k = \{\mathbf{a}_{1..k} | \beta_{k'}^a \leq \beta_{k'} + \varepsilon, \ \forall k', 1 \leq k' \leq k\}, \tag{17}$$

where $\mathbf{a}_{1..k} = [\mathbf{a}_1^T \dots \mathbf{a}_k^T]^T$.

When reasoning about a set of reachable state estimation errors $\mathbf{e}_k^a$ due to stealthy attacks from $\mathcal{A}_k$, we have to also take

---

[7]Formal definition of such policies are presented in the next section.

into account the variability of the estimation error. From (13), we can define a specific region that will contain the error $\mathbf{e}_k^a$ with a desired probability. Therefore, we introduce the following definition.

Definition 2: The k-reachable region $\mathcal{R}_k$ of the state estimation error under the attack (i.e., $\mathbf{e}_k^a$) is the set

$$\mathcal{R}_k = \left\{ \mathbf{e} \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \mathbf{e}\mathbf{e}^T \preccurlyeq E[\mathbf{e}_k^a]E[\mathbf{e}_k^a]^T + \gamma Cov(\mathbf{e}_k^a), \\ \mathbf{e}_k^a = \mathbf{e}_k^a(\mathbf{a}_{1..k}), \; \mathbf{a}_{1..k} \in \mathcal{A}_k \end{array} \right\}. \tag{18}$$

Also, the global reachable region $\mathcal{R}$ of the state estimation error $\mathbf{e}_k^a$ is the set

$$\mathcal{R} = \bigcup_{k=1}^{\infty} \mathcal{R}_k. \tag{19}$$

Here, γ is a design parameter directly related to the desired confidence that $\mathbf{e}_k^a$ belongs to the reachable region. Effectively, the set $\mathcal{R}_k$ captures the set of state estimation errors that can be reached in k^{th} step due to the injected malicious signal, while $\mathcal{R}$ captures the set of all reachable state estimation errors. To assess vulnerability of the system, a critical characteristic of $\mathcal{R}$ is *boundedness* – whether a stealthy attacker can introduce unbounded estimation errors. To simplify the boundedness analysis of $\mathcal{R}$, we start with the following theorem.

Theorem 1: Let $g_k = \mathbf{z}_k^T \mathbf{Q}^{-1} \mathbf{z}_k$ be the detector function. Then, for any $\varepsilon > 0$, such that $\varepsilon \le 1 - \beta_k$, there exists a unique $\alpha > 0$ such that $\beta_k^a \le \beta_k + \varepsilon$ if and only if $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \le \alpha$.

Proof: In the case without attacks, in steady-state g_k has χ^2 distribution with p degrees of freedom, since the residue $\mathbf{z}_k$ is zero-mean ($E[\mathbf{z}_k] = \mathbf{0}$) with covariance matrix $\mathbf{Q} = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T + \mathbf{R}$ [6], [8]. Also, from (10) and (11), $\Delta \mathbf{z}_k = \mathbf{z}_k^a - \mathbf{z}_k$ is output of a deterministic system only controlled by input $\mathbf{a}_{1..k}$, which is assumed to be independent from the value of $\mathbf{z}_k$ (i.e., process and measurement noise) at any particular time k. Thus, $\mathbf{z}_k^a$ is a non-zero mean with covariance matrix $\mathbf{Q}$ – i.e., the attacker is only influencing the $\Delta \mathbf{z}_k = E[\mathbf{z}_k^a - \mathbf{z}_k] = E[\mathbf{z}_k^a]$. Therefore, $\mathbf{g}_k^a = \mathbf{z}_k^{aT} \mathbf{Q}^{-1} \mathbf{z}_k^a$ will have a non-central χ^2 distribution with p degrees of freedom; the non-centrality parameter of this distribution will be $\lambda = \|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}}^2$ [6].

Let h be the threshold for the detector in (7). The alarm probabilities $\beta_k = 1 - P(g_k \le h)$ and $\beta_k^a = 1 - P(g_k^a \le h)$ can be computed from the distributions for g_k and g_k^a as

$$\beta_k = 1 - F_{\chi^2}(h, p), \qquad \beta_k^a = 1 - F_{nc\chi^2}(h; p, \lambda),$$

where $F_{\chi^2}(h, p)$ and $F_{nc\chi^2}(h; p, \lambda)$ are cumulative distribution functions of χ^2 and noncentralized χ^2, respectfully, at h, with p degrees of freedom and noncentrality parameter λ. Since p and h are fixed by the system design, it follows that β_k will be a constant, and β_k^a will be a function of λ.

Consider $\varepsilon = \beta_k^a - \beta_k$. This means that

$$\varepsilon = 1 - F_{nc\chi^2}(h; p, \lambda) - \beta_k. \tag{20}$$

The probability distribution function of non-central χ^2 distribution is smooth (thus making $F_{nc\chi^2}(h; p, \lambda)$ smooth), and $F_{nc\chi^2}(h; p, \lambda)$ is a decreasing function of λ [6]. Hence, it follows that for any ε there will exist exactly one $\sqrt{\lambda} = \|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} = \alpha$ such that (20) is satisfied. Furthermore, for any ε' that is lower than ε, the corresponding $\sqrt{\lambda'} = $

$\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}}$ from (20) has to be lower than α, and vice versa, which concludes the proof. ∎

Since the bound α for $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}}$ in Theorem 1 depends on ε, h and the fact that the χ^2 detector with p degrees of freedom is used, we will denote such value as $\alpha = \alpha_{\chi^2}(\varepsilon, p, h)$.

Remark 2: Related results from [20], [23], focus only on the detection function $g_k = \mathbf{z}_k^T \mathbf{Q}^{-1} \mathbf{z}_k$ and show only sufficient conditions for stealthy attacks – i.e., that in this case from a robustness condition $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \le \alpha$ it follows that the the stealthiness condition $\beta_k^a \le \beta + \varepsilon$ is satisfied. However, the equivalence between conditions $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \le \alpha$ and $\beta_k^a \le \beta + \varepsilon$ will enable us to reduce conservativeness of our analysis as well as analyze boundness of the reachability region for the general type of detection functions from (15), by allowing us to employ both conditions interchangeably.

From Definition 1 and Theorem 1 the following result holds.

Corollary 1: For detection function $g_k = \mathbf{z}_k^T \mathbf{Q}^{-1} \mathbf{z}_k$, there exists $\alpha > 0$ such that the set of all stealthy attacks satisfies

$$\mathcal{A}_k = \{\mathbf{a}_{1..k} | \|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \le \alpha, \; \forall k', 1 \le k' \le k\}. \tag{21}$$

The previous results introduce an equivalent 'robustness-based' representation for the set of stealthy attacks in systems where χ^2 detectors are used. They also provide a foundation to consider the more general formulation (15) for the detector function. We start with the following results characterizing over- and under-approximations of the set $\mathcal{A}_k$ in such case, also using suitable 'robustness-based' representations of the stealthiness condition. By showing that reachable estimation error regions are bounded for these sets of attacks, we will be able to reason whether the reachable region of state estimation errors is bounded for attacks from the set $\mathcal{A}_k$.

Lemma 1: For a system with the detector function g_k^a of the form from (15), the set of all stealthy attacks $\mathcal{A}_k$ can be underapproximated by the set

$$\underline{\mathcal{A}}_k = \{\mathbf{a}_{1..k} | \|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \le \underline{\alpha}, \; \forall k', 1 \le k' \le k\} \tag{22}$$

(i.e., $\underline{\mathcal{A}}_k \subseteq \mathcal{A}_k$), where $\underline{\alpha} = \alpha_{\chi^2}(\varepsilon, \mathcal{T}p, h/c_{max})/\sqrt{\mathcal{T}}$.

In essence, the lemma states that if $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \le \underline{\alpha}$ holds, then $g_k^a \le h$ for the general detection function from (15) is satisfied with probability that is lower than or equal to $\beta_k + \varepsilon$.

Proof: Consider an attack sequence $\mathbf{a}_{1..k} \in \underline{\mathcal{A}}_k$ and the resulting evolution of the system from (10) and (11), with $\|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \le \underline{\alpha}$, for all $k', 1 \le k' \le k$. Then,

$$\sum_{i=k-\mathcal{T}+1}^{k} \|\Delta \mathbf{z}_i\|_{\mathbf{Q}^{-1}}^2 \le \sum_{i=k-\mathcal{T}+1}^{k} \underline{\alpha}^2 = \alpha_{\chi^2}^2(\varepsilon, \mathcal{T}p, h). \tag{23}$$

In addition, we define $c_{max} = \max(c_1, \ldots, c_{\mathcal{T}})$ and

$$\underline{g}_k^a = \sum_{i=k-\mathcal{T}+1}^{k} c_{max} \mathbf{z}_i^{aT} \mathbf{Q}^{-1} \mathbf{z}_i^a, \tag{24}$$

as well as $\underline{\beta_k^a} = P(\underline{g}_k^a > h)$. From (24), $\underline{g}_k^a$ is a scaled sum of noncentral χ^2 distributions with p degrees of freedom, so $\underline{g}_k^a/c_{max}$ will have the noncentral χ^2 distribution with $p\mathcal{T}$ degrees of freedom and the central moment equal to

$$\underline{\lambda} = \sum_{i=k-\mathcal{T}+1}^{k} \|\Delta \mathbf{z}_i\|_{\mathbf{Q}^{-1}}^2. \tag{25}$$

Since $\underline{\beta_k^a} = P(\underline{g_k^a} > h) = P(\underline{g_k^a}/c_{max} > h/c_{max})$, following the proof of Theorem 1 for the $\underline{g_k^a}/c_{max}$ detection function we have that $\underline{\beta_k^a} \leq \beta_k + \varepsilon$ is satisfied if and only if $\sqrt{\underline{\lambda}} \leq \alpha_{\chi^2}(\varepsilon, p\mathcal{T}, h/c_{max})$. That is, using (25)

$$\left(\underline{\beta_k^a} \leq \beta_k + \varepsilon\right) \Leftrightarrow \sum_{i=k-\mathcal{T}+1}^{k} \|\Delta \mathbf{z}_i\|_{\mathbf{Q}^{-1}}^2 \leq \alpha_{\chi^2}^2(\varepsilon, p\mathcal{T}, h/c_{max}). \tag{26}$$

Since (23) follows from the condition of the theorem, from (26) we have that $\underline{\beta_k^a} \leq \beta_k + \varepsilon$ is satisfied. From (24) we have that $g_k^a \leq \underline{g_k^a}$, meaning that $\beta_k^a \leq \underline{\beta_k^a}$. Thus, $\left(\beta_k^a \leq \beta_k + \varepsilon\right)$ holds, and $\mathbf{a}_{1..k} \in \mathcal{A}_k$ (i.e., $\underline{\mathcal{A}_k} \subseteq \mathcal{A}_k$). ∎

Lemma 2: For a system with the detector function g_k^a of the form from (15), the set of all stealthy attacks at time k, $\mathcal{A}_k$, can be overapproximated by the set

$$\overline{\mathcal{A}}_k = \{\mathbf{a}_{1..k} | \|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \leq \overline{\alpha}, \ \forall k', 1 \leq k' \leq k\}, \tag{27}$$

(i.e., $\overline{\mathcal{A}}_k \supseteq \mathcal{A}_k$), where $\overline{\alpha} = \alpha_{\chi^2}(\varepsilon, p, h/c_{\mathcal{T}})$.

Proof: Consider an attack sequence $\mathbf{a}_{1..k} \in \mathcal{A}_k$ with the detector function g_k^a from (15). Let $\overline{g_k^a} = c_{\mathcal{T}} \mathbf{z}_k^{aT} \mathbf{Q}^{-1} \mathbf{z}_k^a$. Since $\overline{g_k^a} \leq g_k^a$ it follows that $\overline{\beta_k^a} = P(\overline{g_k^a} > h) \leq \beta_k^a$, where β_k^a is defined as in (14). Since $\mathbf{a}_{1..k}$ are stealthy, it follows that $\beta_k^a \leq \beta_k + \varepsilon$, and thus $\overline{\beta_k^a} \leq \beta_k + \varepsilon$ holds.

On the other hand, the function $\overline{g_k^a}/c_{\mathcal{T}}$ has the χ^2 distribution; by following the proof of Theorem 1 for $\overline{g_k^a}/c_{\mathcal{T}}$ we have that $\overline{\beta_k^a} \leq \beta_k + \varepsilon$ is satisfied if and only if $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \leq \overline{\alpha} = \alpha_{\chi^2}(\varepsilon, p, h/c_{\mathcal{T}})$. Therefore, we have that $\beta_{k'}^a \leq \beta_{k'} + \varepsilon$ implies $\|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \leq \overline{\alpha}$, $k' = 1, ..., k$, meaning that $\mathbf{a}_{1..k} \in \overline{\mathcal{A}}_k$ (i.e., $\mathcal{A}_k \subseteq \overline{\mathcal{A}}_k$). ∎

Remark 3: The previous lemmas also hold for the detection function $g_k = \sum_{i=1}^{k} c_k \mathbf{z}_i^T \mathbf{Q}^{-1} \mathbf{z}_i$; this can be shown by replacing $\mathcal{T}$ with k in the previous analysis, since it would not affect their proofs. This means that these results hold for both windowed detectors as well as the detectors previously used in security-related scenarios in [11], [12], [30] – the use of such detectors is explored in detail in Section V.

Lemmas 1 and 2 introduce attack sets $\underline{\mathcal{A}}_k$ and $\overline{\mathcal{A}}_k$ for which the attack constraints are captured as robustness bounds on $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}}$ instead of probabilities of attack detection, and for which $\underline{\mathcal{A}}_k \subseteq \mathcal{A}_k \subseteq \overline{\mathcal{A}}_k$. Hence, to analyze impact of stealthy attacks, we can consider the effects of attacks that have to maintain $\|\Delta \mathbf{z}_k\|$ below a certain threshold.

Theorem 2: $\mathcal{R}_k$ from (18) is bounded if and only if the set

$$\hat{\mathcal{R}}_k^{\alpha} = \left\{ \Delta \mathbf{e}_k \in \mathbb{R}^n \ \middle| \ \begin{array}{c} \Delta \mathbf{e}_k, \Delta \mathbf{z}_k \text{ satisfy (10) and (11),} \\ \Delta \mathbf{e}_{k-1} \in \hat{\mathcal{R}}_{k-1}^{\alpha}, \|\Delta \mathbf{z}_k\|_2 \leq \alpha \end{array} \right\}, \tag{28}$$

is bounded, where $\hat{\mathcal{R}}_0^{\alpha} = \mathbf{0} \in \mathbb{R}^n$ and $\alpha > 0$.

Proof: From (13), $\gamma Cov(\mathbf{e}_k^a)$ is bounded and we can simplify our presentation by focusing on the case where $\gamma = 0$. Furthermore, for any vector $\mathbf{v}$, the set $\{\mathbf{e} \in \mathbb{R}^n \mid \mathbf{e}\mathbf{e}^T \preccurlyeq \mathbf{v}\mathbf{v}^T\}$ is bounded if and only if the vector $\mathbf{v}$ is bounded. Therefore, the set $\{\mathbf{e} \in \mathbb{R}^n \mid \mathbf{e}\mathbf{e}^T \preccurlyeq E[\mathbf{e}_k^a]E[\mathbf{e}_k^a]^T + \gamma Cov(\mathbf{e}_k^a)\}$ will be bounded if and only if $E[\mathbf{e}_k^a] = \Delta \mathbf{e}_k$ (from (12)) is bounded.

Consider attack vectors $\mathbf{a}_{1..k} \in \mathcal{A}_k$. From Lemmas 1 and 2 we have that

$$\{\Delta \mathbf{e}_k | \mathbf{a}_{1..k} \in \underline{\mathcal{A}}_k\} \subseteq \{\Delta \mathbf{e}_k | \mathbf{a}_{1..k} \in \mathcal{A}_k\} \subseteq \{\Delta \mathbf{e}_k | \mathbf{a}_{1..k} \in \overline{\mathcal{A}}_k\}, \tag{29}$$

where we somewhat abuse the notation, by having $\{\Delta \mathbf{e}_k | \mathbf{a}_{1..k} \in \mathcal{A}\}$ capture all reachable vectors $\Delta \mathbf{e}_k$ when the system (10) is 'driven' by attack vectors from the set $\mathcal{A}$. On the other hand, from linearity of the system described by (10) and (11), the sets $\{\Delta \mathbf{e}_k | \|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \leq \underline{\alpha}, k' = 1, ..., k\}$ and $\{\Delta \mathbf{e}_k | \|\Delta \mathbf{z}_{k'}\|_{\mathbf{Q}^{-1}} \leq \overline{\alpha}, k' = 1, ..., k\}$ are either both bounded or both unbounded. Thus, from (29), these sets are bounded if and only if $\{\Delta \mathbf{e}_k | \mathbf{a}_{1..k} \in \mathcal{A}_k\}$ is bounded.

Finally, as $\frac{1}{|\lambda_{max}|} \|\Delta \mathbf{z}_k\|_2 \leq \|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \leq \frac{1}{|\lambda_{min}|} \|\Delta \mathbf{z}_k\|_2$, where $\lambda_{max}, \lambda_{min}$ are the largest and smallest, respectively, eigenvalue of $\mathbf{Q}$, the region $\hat{\mathcal{R}}_k^{\alpha}$ will be bounded for the constraint $\|\Delta \mathbf{z}_k\|_{\mathbf{Q}^{-1}} \leq \alpha$ if and only if its bounded with a 2-norm stealthiness constraint $\|\Delta \mathbf{z}_k\|_2 \leq \alpha$ from (28). ∎

A. Perfectly Attackable Systems

Theorem 2 can be used to formally capture dynamical systems for which there exists a stealthy attack sequence that results in an unbounded state estimation error – i.e., for such systems, given enough time, the attacker can make arbitrary changes in the system states without risking detection.

Definition 3: A system is *perfectly attackable* (PA) if the system's reachable set $\mathcal{R}$ from (19) is an unbounded set.

As shown in [11], [22], for LTI systems without any additional data integrity guarantees, the set $\hat{\mathcal{R}}^{\alpha} = \bigcup_{k=0}^{\infty} \hat{\mathcal{R}}_k^{\alpha}$ can be bounded or unbounded depending on the system dynamics and the set of compromised sensors $\mathcal{K}$. From Theorem 2, this property is preserved for the set $\mathcal{R}$ as well. For this reason, we will be using the definition of $\hat{\mathcal{R}}^{\alpha}$ to analyze boundedness of $\mathcal{R}$, and to simplify the notation due to linearity of the constraint we will assume that $\alpha = 1$ – i.e., for this analysis we consider the stealthiness attack constraint as

$$\|\Delta \mathbf{z}_k\|_2 \leq 1, \qquad k \in \mathbb{N}_0, \tag{30}$$

imposed on the system Ξ from (10) and (11).

Now, the theorem below follows from [11], [22].

Theorem 3: A system from (9) is perfectly attackable if and only if the matrix $\mathbf{A}$ is unstable, and at least one eigenvector $\mathbf{v}$ corresponding to an unstable eigenvalue satisfies $\text{supp}(\mathbf{Cv}) \subseteq \mathcal{K}$ and $\mathbf{v}$ is a reachable state of the system Ξ from (10), (11).

Note that [22] also uses the term *unstable eigenvalue* λ to denote $|\lambda| \geq 1$. In the next section, we show that intermittent integrity guarantees significantly limit stealthy attacks even for *perfectly-attackable* systems.

IV. STEALTHY ATTACKS IN SYSTEMS WITH INTERMITTENT INTEGRITY ENFORCEMENT

In this section, we analyze the effects that intermittent data integrity guarantees have on the estimation error under attack. To formalize this notion, we start with the following definition.

Definition 4: A *global* intermittent data integrity enforcement policy (μ, f, L), where $\mu = \{t_k\}_{k=0}^{\infty}$ such that $t_0 > 1$,

for all $k > 0$, $t_{k-1} < t_k$ and $L = \sup_{k>0} (t_k - t_{k-1})$, ensures that

$$\mathbf{a}_{t_k} = \mathbf{a}_{t_k+1} = ... = \mathbf{a}_{t_k+f-1} = \mathbf{0}, \forall k \geq 0.$$

Furthermore, for a sensor $s_i \in \mathcal{S}$, the sensor's intermittent data integrity enforcement policy (μ_i, f_i, L_i), where $\mu_i = \{t_k^i\}_{k=0}^{\infty}$ with $t_0^i > 1$, $t_{k-1}^i < t_k^i$ for all $k > 0$, and $L_i = \sup_{k>0} \left(t_k^i - t_{k-1}^i \right)$, ensures that

$$\mathbf{a}_{t_k^i,i} = \mathbf{a}_{t_k^i+1,i} = ... = \mathbf{a}_{t_k^i+f_i-1,i} = 0, \forall k \geq 0.$$

Intuitively, an intermittent data integrity enforcement policy for sensor s_i ensures that the injected attack $\mathbf{a}_{k,i}$ via the sensor will be equal to zero in at least f_i consecutive points, where the starts of these 'blocks' are at most L_i time-steps apart. Similarly, for a *global* intermittent data integrity enforcement policy, the whole attack vector $\mathbf{a}_k$ has to be $\mathbf{0}$ for at least f consecutive steps, and the duration between these blocks is bounded from above to at most L time-steps.

Global intermittent integrity enforcement is easier to model (and analyze, as we will show in the next section). However, compared to the use of p separate sensor's intermittent integrity enforcements, global enforcement policies impose significantly larger communication and computation overhead in every time-step when data integrity is enforced. For example, with global enforcement every sensor has to be able to compute and add a MAC to its message transmitted over a shared bus during one sampling period (which usually corresponds to a single communication frame). In addition, since in these systems estimation/control updates are commonly computed once all messages are received, when the integrity is enforced the estimator has to be able to evaluate/recompute all received MACs before its execution for that time-period. On the other hand, with integrity enforcement for each sensor, their MACs can be sent and reevaluated in separate (e.g., consecutive) sampling periods (i.e., communication frames).

Remark 4: It is worth noting that our definition of intermittent integrity enforcement policies imposes a maximum time between integrity enforcements which, as we will show, is related to the worst-case estimation error caused by the attacks. The definition also captures **periodic** integrity enforcements when $L = t_k - t_{k-1}$ for all $k > 0$. Finally, the definition also allows for capturing policies with continuous integrity enforcements, by specifying $L \leq f$.

The following theorem specifies that when a global intermittent integrity enforcement policy is used a stealthy attacker cannot insert an unbounded expected state estimation error.

Theorem 4: Consider an LTI system from (1) with a global data integrity policy (μ, f, L), where

$$f = \min(\psi, q_{un}), \tag{31}$$

L is finite, ψ is the observability index of the $(\mathbf{A}, \mathbf{C})$ pair, and q_{un} denotes the number of unstable eigenvalues of $\mathbf{A}$. Then the system is not perfectly attackable.

From the above theorem, it follows that even intermittent integrity guarantees significantly limit the damage that the attacker could make to the system. Furthermore, the theorem makes no assumptions about the set $\mathcal{K}$ of compromised sensors; in the general case, system designers may not be able

to provide this type of guarantees during system design, and thus no restrictions are imposed on the set, neither regarding the number of elements or whether some sensors belong to it.

Remark 5: In our preliminary results reported in [7], a similar formulation of Theorem 4 is used with $f = \min \left(nullity(\mathbf{C}) + 1, q_{un} \right)$. Since $\psi \leq n - rank(\mathbf{C}) + 1$ from [36], using the rank–nullity theorem it follows that $\psi \leq nullity(\mathbf{C}) + 1$, meaning that the condition from Theorem 4 is stronger than our earlier result and may further reduce the number of integrity-enforcement points.

In the rest of the paper, we use the notation from Theorem 4 for f and q_{un}. To show the theorem, we exploit the following Lemma 3 and Theorem 5; the lemma states that if stealthy attacks introduce unbounded estimation error $\Delta \mathbf{e}_k$, the unbounded components must belong to vector subspaces corresponding to unstable modes of the system (i.e., matrix $\mathbf{A}$).

Lemma 3: Consider system Ξ from (10) and (11) under the stealthiness constraint (30), and let us denote by $\mathbf{v}_1, \ldots, \mathbf{v}_{q_{un}}$ eigenvectors and generalized eigenvectors that correspond to unstable eigenvalues of matrix $\mathbf{A}$. Then, unbounded estimation errors $\Delta \mathbf{e}_k$ can be represented as

$$\Delta \mathbf{e}_k = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_{q_{un}} \mathbf{v}_{q_{un}} + \varrho_k \tag{32}$$

where $\varrho_k = \sum_{j=q_{un}+1}^{n} \alpha_j \mathbf{v}_j$ is a bounded vector, and for some $1 \leq i \leq q_{un}$ it holds that $\alpha_i \to \infty$ as $k \to \infty$.

Proof: The proof is provided in the appendix. ∎

Theorem 5: Consider any $k \in \mathbf{N}$, such that $k + 1 \in \mu$ (i.e., at time $k + 1$ an integrity enforcement block in the policy μ starts). If $\Delta \mathbf{e}_k$ is reachable state of Ξ, and if vectors $\mathbf{CA}\Delta \mathbf{e}_k, \mathbf{CA}\Delta \mathbf{e}_{k+1}, \ldots, \mathbf{CA}\Delta \mathbf{e}_{k+f-1}$ are bounded, then the vector $\Delta \mathbf{e}_{k+f}$ has to be bounded for any stealthy attack.[8]

Proof: From (10) and (11) it follows that

$$\Delta \mathbf{e}_{k+f} = \mathbf{A}\Delta \mathbf{e}_{k+f-1} - \mathbf{K}\Delta \mathbf{z}_{k+f}, \tag{33}$$

$$\Delta \mathbf{z}_{k+f} = \mathbf{CA}\Delta \mathbf{e}_{k+f-1} + \mathbf{a}_{k+f}. \tag{34}$$

Since $\|\mathbf{K}\|_2$ is bounded, and $\|\Delta \mathbf{z}_{k+f}\|_2 \leq 1$ due to the stealthy attack constraint (30), then $\|\mathbf{K}\Delta \mathbf{z}_{k+f}\|_2 \leq \|\mathbf{K}\|_2 \|\Delta \mathbf{z}_{k+f}\|_2$ is bounded. Thus, to show that $\|\Delta \mathbf{e}_{k+f}\|_2$ is bounded, it is sufficient to prove that $\|\mathbf{A}\Delta \mathbf{e}_{k+f-1}\|_2$ is bounded.

Let's assume the opposite – i.e., that $\|\mathbf{A}\Delta \mathbf{e}_{k+f-1}\|_2$ is unbounded while $\|\mathbf{CA}\Delta \mathbf{e}_k\|_2, \ldots, \|\mathbf{CA}\Delta \mathbf{e}_{k+f-1}\|_2$ are all bounded. From (33) it follows that

$$\mathbf{A}\Delta \mathbf{e}_{k+f-1} = \mathbf{A}^f \Delta \mathbf{e}_k - \sum_{j=1}^{f-1} \mathbf{A}^{f-j} \mathbf{K}\Delta \mathbf{z}_{k+j}.$$

Given that $\|\Delta \mathbf{z}_{k+1}\|_2, \ldots, \|\Delta \mathbf{z}_{k+f-1}\|_2$ are bounded due to the stealthy attack requirements, in order for $\mathbf{A}\Delta \mathbf{e}_{k+f-1}$ to be unbounded, $\mathbf{A}^f \Delta \mathbf{e}_k$ has to be unbounded as well.

As $\mathbf{CA}\Delta \mathbf{e}_{k+1}$ is bounded, this implies that $\mathbf{CA}(\mathbf{A}\Delta \mathbf{e}_k - \mathbf{K}\Delta \mathbf{z}_{k+1})$ has to be bounded too. However, as $\Delta \mathbf{z}_{k+1}$ has to be bounded due to the stealthiness condition, it follows that

---

[8]Formally, the theorem states that the subsequence $\{\Delta \mathbf{e}_{k+f}\}_{(k+1)\in\mu}$ of the sequence $\{\Delta \mathbf{e}_k\}_{k\in\mathbb{N}}$ is bounded, if the subsequence $\{\mathbf{CA}\Delta \mathbf{e}_k, \mathbf{CA}\Delta \mathbf{e}_{k+1}, \ldots, \mathbf{CA}\Delta \mathbf{e}_{k+f-1}\}_{(k+1)\in\mu}$ of the sequence $\{\mathbf{CA}\Delta \mathbf{e}_k\}_{k\in\mathbb{N}}$ is bounded. However, to simplify our presentation and notation, we simply refer to the vectors, instead of subsequences, as bounded.

$\mathbf{CA}^2\Delta\mathbf{e}_k$ has to remain bounded. Similarly, we can show that this holds up to $\mathbf{CA}^f\Delta\mathbf{e}_k$; thus the vector $\mathbf{b}_k(f)$ defined as

$$\mathbf{b}_k(f) = \underbrace{\begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \dots \\ \mathbf{CA}^{f-1} \end{bmatrix}}_{\mathcal{O}_f} \mathbf{A}\Delta\mathbf{e}_k \qquad (35)$$

is bounded. Now, we consider two cases.

Case I: If f is observability index of $(\mathbf{A}, \mathbf{C})$ pair (i.e., $f = \psi$), then $\mathcal{O}_f$ has full rank, from which it follows that $\mathbf{A}\Delta\mathbf{e}_k$ (and thus $\mathbf{A}^f\Delta\mathbf{e}_k$) has to be also bounded, which is a contradiction.

Case II: Consider $f = q_{un}$, and let us use similarity transform $\mathbf{V}$ on the initial system, where $\mathbf{V}$ is defined as in the Lemma 3 proof – i.e., $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_n]$ and we index (generalized) eigenvectors such that for each eigenvector $\mathbf{v}_i$ with L_i generalized eigenvectors, $\mathbf{v}_{i+1}, ..., \mathbf{v}_{i+L_i}$ is its generalized eigenvector chain; in addition, $\mathbf{v}_1, ..., \mathbf{v}_{q_{un}}$ are the eigenvectors (including generalized eigenvectors) for all unstable modes of $\mathbf{A}$.

Thus, the transformed system can be captured as

$$\tilde{\mathbf{A}} = \mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{J} = \begin{bmatrix} \mathbf{J}_{1(q_{un}\times q_{un})} & \mathbf{0}_{(q_{un}\times n - q_{un})} \\ \mathbf{0}_{(n - q_{un}\times q_{un})} & \mathbf{J}_{2(n - q_{un}\times n - q_{un})} \end{bmatrix} \quad (36)$$

$$\tilde{\mathbf{C}} = \mathbf{CV} = \begin{bmatrix} \tilde{\mathbf{C}}_{1(p\times q_{un})} & \tilde{\mathbf{C}}_{2(p\times n - q_{un})} \end{bmatrix}$$

where $\mathbf{J}$ is the Jordan form of $\mathbf{A}$, $\mathbf{J}_1$ captures unstable modes of $\mathbf{A}$ and the pair $(\tilde{\mathbf{A}}, \tilde{\mathbf{C}})$ is also observable.

Since $\mathbf{A}^f\Delta\mathbf{e}_k$ is unbounded we have that $\Delta\mathbf{e}_k$ is unbounded (from $\|\mathbf{A}^f\Delta\mathbf{e}_k\|_2 \le \|\mathbf{A}\|_2^f\|\Delta\mathbf{e}_k\|_2$). Thus, from Lemma 3,

$$\Delta\mathbf{e}_k = \mathbf{V}[\alpha_1 \dots \alpha_n]^T = \mathbf{V}\alpha_{1..n}, \qquad (37)$$

where $\alpha_{1..q_{un}} = \begin{bmatrix} \alpha_1 \dots \alpha_{q_{un}} \end{bmatrix}^T$ is unbounded while $\alpha_{(q_{un}+1)..n} = \begin{bmatrix} \alpha_{(q_{un}+1)} \dots \alpha_{q_n} \end{bmatrix}^T$ is a bounded vector. Since $\tilde{\mathbf{C}}\tilde{\mathbf{A}}^j = \tilde{\mathbf{C}}\mathbf{J}^j = \begin{bmatrix} \tilde{\mathbf{C}}_1\mathbf{J}_1^j & \tilde{\mathbf{C}}_2\mathbf{J}_2^j \end{bmatrix}$, from (35) it follows that

$$\mathbf{b}_k(f) = \begin{bmatrix} \tilde{\mathbf{C}}_1 & \tilde{\mathbf{C}}_2 \\ \tilde{\mathbf{C}}_1\mathbf{J}_1 & \tilde{\mathbf{C}}_2\mathbf{J}_2 \\ \dots & \\ \tilde{\mathbf{C}}_1\mathbf{J}_1^{f-1} & \tilde{\mathbf{C}}_2\mathbf{J}_2^{f-1} \end{bmatrix} \mathbf{J}\alpha_{1..n} =$$

$$= \underbrace{\begin{bmatrix} \tilde{\mathbf{C}}_1 \\ \tilde{\mathbf{C}}_1\mathbf{J}_1 \\ \dots \\ \tilde{\mathbf{C}}_1\mathbf{J}_1^{f-1} \end{bmatrix}}_{\tilde{\mathcal{O}}_{uns,f}} \mathbf{J}_1\alpha_{1..q_{un}} + \underbrace{\begin{bmatrix} \tilde{\mathbf{C}}_2 \\ \tilde{\mathbf{C}}_2\mathbf{J}_2 \\ \dots \\ \tilde{\mathbf{C}}_2\mathbf{J}_2^{f-1} \end{bmatrix}}_{\tilde{\mathcal{O}}_{sta,f}} \mathbf{J}_2\alpha_{(q_{un}+1)..n}. \quad (38)$$

Also, as $\mathbf{b}_k(f)$ and $\alpha_{(q_{un}+1)..n}$ are bounded, from (38) vector

$$\tilde{\mathbf{b}}_k(f) = \tilde{\mathcal{O}}_{uns,f}\mathbf{J}_1\alpha_{1..q_{un}} \qquad (39)$$

is also bounded. Note that $\tilde{\mathcal{O}}_{uns,f}$ is effectively the observability matrix of the $(\mathbf{J}_1, \tilde{\mathbf{C}}_1)$ pair corresponding to the subsystem with the q_{un} unstable eigenvalues of $\mathbf{A}$.

To show that $(\mathbf{J}_1, \tilde{\mathbf{C}}_1)$ is observable, let us assume the opposite; thus, there exist an eigenvector $\tilde{\mathbf{v}}_j$ of $\mathbf{J}_1$ such that $\tilde{\mathbf{v}}_j \in null(\tilde{\mathbf{C}}_1) \subseteq \mathbb{R}^{q_{un}}$. Take note that $\mathbf{J}_1$ is a Jordan matrix,
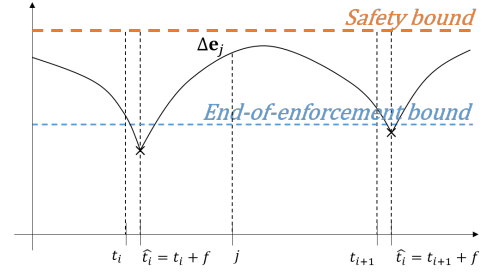

Fig. 3. Estimation error evolution between consecutive integrity enforcement blocks.

so each of its eigenvectors has to be a projection vector $\mathbf{i}_j$ (as defined in Sec. I-A), where j, $1 \le j \le q_{un}$, corresponds to the start of a Jordan block of $\mathbf{J}_1$. Yet this implies that $\tilde{\mathbf{C}}_1\mathbf{i}_j = \mathbf{0}_{p\times 1}$ – i.e., the jth column of $\tilde{\mathbf{C}}_1$ and thus the jth column of $\tilde{\mathbf{C}}$ are zero vectors. However, since $\tilde{\mathbf{C}} = \mathbf{CV} = [\mathbf{Cv}_1 \dots \mathbf{Cv}_n]$, it follows that $\mathbf{Cv}_j = \mathbf{0}$ for some j. Due to the way $\mathbf{V}$ is formed and since j has to be the start index of a Jordan block in $\mathbf{J}_1$, it follows that $\mathbf{v}_j$ is an eigenvector of $\mathbf{A}$. However, this implies that $\mathbf{v}_j \in null(\mathbf{C})$, making $(\mathbf{A}, \mathbf{C})$ pair unobservable and contradicting our initial assumption about the system.

Therefore, $(\mathbf{J}_1, \tilde{\mathbf{C}}_1)$ is observable, meaning that $\tilde{\mathcal{O}}_{uns,f}$ is full rank. Furthermore, $\mathbf{J}_1$ is invertible as it contains only unstable (i.e., non-zero) eigenvalues of the system on the diagonal. Hence, from (39) and the fact that $\mathbf{b}_k(f)$ is bounded, it follows that vector $\alpha_{1..q_{un}}$ has to be bounded, which from (37) contradicts that $\Delta\mathbf{e}_k$ and $\mathbf{A}^f\Delta\mathbf{e}_k$ are unbounded, and thus concludes the proof. ∎

Using the previous theorem, we now prove Theorem 4.

Proof: [Proof of Theorem 4] Consider any time-point $t_k + f$ such that $t_k \in \mu$ – i.e., t_k is the start of an integrity enforcement block. Thus, $\mathbf{a}_{t_k} = ... = \mathbf{a}_{t_k+f-1} = \mathbf{0}$. From (11) it follows that $\Delta\mathbf{z}_{t_k+j} = \mathbf{CA}\Delta\mathbf{e}_{t_k+j-1}$, $j = 0, ..., f-1$, and thus from (30)

$$\|\mathbf{CA}\Delta\mathbf{e}_{t_k+j-1}\|_2 \le 1, \qquad j = 0, ..., f-1.$$

Now, from Theorem 5 it follows that the state estimation error $\Delta\mathbf{e}_{t_k+f-1}$ has to be bounded for any stealthy attack; this holds for all time points at the ends of integrity enforcement intervals. Since in the proof of Theorem 5, we have not used any specifics of the time points, there exists a global bound on state estimation error at the end of all integrity enforcement periods (as illustrated in Figure 3).

Finally, consider the expected state-estimation error vector at any time j. From Definition 4, there exists $t_i \in \mu$ such that $j \in \left[\hat{t}_i, \hat{t}_i + L\right)$, and $\hat{t}_i = t_i + f$ (Figure 3). From (10) and (11), it follows that

$$\Delta\mathbf{e}_j = \mathbf{A}^{j-\hat{t}_i}\Delta\mathbf{e}_{\hat{t}_i} - \sum_{l=1}^{j-\hat{t}_i} \mathbf{A}^{j-\hat{t}_i-l}\mathbf{K}\Delta\mathbf{z}_{\hat{t}_i+l}. \qquad (40)$$

Thus, the evolution of the expected state estimation error vector between two time points with bounded values can be described as evolution over a finite number of steps of a dynamical system with bounded inputs (since $\|\Delta\mathbf{z}_{\hat{t}_i+l}\|_2 \le 1$);

from the triangle and Cauchy-Schwarz inequalities it follows

$$\|\Delta \mathbf{e}_j\|_2 \leq \|\mathbf{A}\|_2^{j-\hat{t}_i} \|\Delta \mathbf{e}_{\hat{t}_i}\|_2 + \sum_{l=1}^{j-\hat{t}_i} \|\mathbf{A}\|_2^{j-\hat{t}_i-l} \|\mathbf{K}\|_2. \quad (41)$$

Hence, the expected estimation error vector $\Delta \mathbf{e}_j$ is bounded for any j, and the system is not perfectly attackable. ∎

From Theorem 5 proof, a stronger condition for f holds.

Corollary 2: Consider an LTI system from (1) with a global data integrity policy (μ, f, L), where L is finite and f is the observability index of $(\mathbf{J}_1, \check{\mathbf{C}}_1)$ pair, with $\mathbf{J}_1$ and $\check{\mathbf{C}}_1$ defined as in (36). Then the system is not perfectly attackable.

Proof: Follows directly from Theorem 4 and 5 proofs. ∎

Theorem 4 assumes that the attacker has the full knowledge of the system's integrity enforcement policy – i.e., at which time-points integrity enforcements will occur. As we illustrate in Section VI, this allows the attacker to plan attacks that maximize the error, while ensuring stealthiness of the attack by reducing the state estimation errors to the levels that will not trigger detection during integrity enforcement intervals. On the other hand, if the attacker does not have the knowledge of μ (i.e., if (s)he is not aware of the time points in which integrity enforcements would occur), the integrity enforcement requirements can be additionally relaxed; as the attacker does not know when enforcements occur, (s)he has to ensure that if at any future point (including the next time-step) malicious data cannot be injected, the residue would still remain below the threshold (30). Thus, we obtain the following result.

Theorem 6: Any LTI system from (1) with a global data integrity policy $(\mu, 1, L)$ (i.e., with $f = 1$) is not perfectly attackable for any stealthy attacker that does not know the time points μ when data integrity is enforced.

Proof: First, note that the sequence $\mathbf{CA}\Delta \mathbf{e}_k$ cannot be bounded if the attacker wants to introduce an unbounded state estimation error. If it was bounded, from (11) and (30) it would follow that $\mathbf{a}_k$ is always bounded; this in turn would imply that the system from (10) has bounded inputs, which since matrix $(\mathbf{A} - \mathbf{KCA})$ is stable ($\mathbf{K}$ is Kalman gain) would imply that $\Delta \mathbf{e}_k$ cannot diverge – the reachable set $\mathcal{R}$ can not be unbounded.

On the other hand, let us assume that the system is perfectly attackable – i.e., the expected state estimation error can be unbounded. Then, from our previous argument it follows that $\mathbf{CA}\Delta \mathbf{e}_k$ is unbounded and thus we can find k and $\Delta \mathbf{e}_k$ such that $\|\mathbf{CA}\Delta \mathbf{e}_k\|_2 > 1$. Then, if global data integrity is enforced only once at the time-step $k+1$, from (11) it would follow that $\|\Delta \mathbf{z}_{k+1}\| = \|\mathbf{CA}\Delta \mathbf{e}_k\|_2 > 1$, which violates the stealthiness requirement from (30). ∎

Theorems 4 and 5 consider a worst-case scenario without any constraints or assumptions about the set of compromised sensors $\mathcal{K}$ (e.g., that less than q sensors are compromised). Yet, some knowledge about the set $\mathcal{K}$ may be available at design-time. For instance, for *MitM* attacks some sensors cannot be in set $\mathcal{K}$, such as on-board sensors that do not communicate over a network to deliver information to the estimator, or sensors with built-in continuous data authentication. In these cases, the number of integrity enforcements can be reduced.

Corollary 3: Consider a system from (1) with a global data integrity policy (μ, f, L), where $f = \min(\psi, q_{un}^*)$, ψ is the observability index of $(\mathbf{A}, \mathbf{C})$, and q_{un}^* denotes the number of unstable eigenvalues λ_i of $\mathbf{A}$ for which the corresponding eigenvector $\mathbf{v}_i$ satisfies $\text{supp}(\mathbf{Cv}_i) \in \mathcal{K}$. Then the system is not perfectly attackable.

Proof: The proof directly follows the proof of Theorem 4, with the only difference that all $\alpha_i \to \infty$ from Lemma 3 also have to correspond to the unstable eigenvectors $\mathbf{v}_i$ satisfying that $\text{supp}(\mathbf{Cv}_i) \in \mathcal{K}$; otherwise, consider $\alpha_i \to \infty$, from a decomposition of a 'large' $\Delta \mathbf{e}_k$ such that $\|\mathbf{C}\alpha_i\mathbf{v}_i\|_2 \to \infty$ and $\text{supp}(\mathbf{Cv}_i) \notin \mathcal{K}$. Then the components of the residue $\Delta \mathbf{z}_{k+1}$ whose indicies are in $\text{supp}(\mathbf{Cv}_i)$ but not in $\mathcal{K}$ (i.e., $P_{\text{supp}(\mathbf{Cv}_i)\setminus\mathcal{K}}\Delta \mathbf{z}_{k+1}$) cannot be influenced by the attack signal $\mathbf{a}_{k+1}$, meaning that their large values due to $\alpha_i \to \infty$ cannot be compensated for by the attack signal, and thus will violate the stealthiness condition (30). ∎

Let us recall Definition 1 that introduced $\mathcal{A}_k$, the set of all stealthy attacks up to time k – it only requires that attack vector $\mathbf{a}_{1..k} \in \mathcal{A}_k$ satisfies the stealthiness conditions up to time k. Thus, as shown in the proof of Theorem 4, the attacker applying attack $\mathbf{a}_{1..k} \in \mathcal{A}_k$ may have to violate the stealthiness constraint during the next integrity enforcement block, since for those time-points t when integrity is enforced $\mathbf{a}_t = \mathbf{0}$. As the attacker's goal is to remain stealthy even during integrity enforcements, we consider policy-aware stealthy attack sets.

Definition 5: For an integrity enforcement policy (μ, f, L), the set of all policy-aware stealthy attacks up to time k is

$$\mathcal{A}_k^\mu = \left\{ \mathbf{a}_{1..k} \left| \begin{array}{c} \mathbf{a}_{1..k'} \in \mathcal{A}_{k'}, \\ k' = \min\{t \mid (t - f + 1 \in \mu) \wedge (t \geq k)\} \end{array} \right. \right\}.$$

Intuitively, the attacker will always plan attacks at least until the end of next integrity enforcement block (captured by k'), while keeping the probability of detection as low. Thus, we also need to modify the definition of the k-reachable region $\mathcal{R}_k$ (Def. 2), as it depends on the employed set of stealthy attacks.

Definition 6: The policy aware k-reachable region $\mathcal{R}_k^\mu$ of the state estimation error under the attack (i.e., $\mathbf{e}_k^a$) is the set

$$\mathcal{R}_k^\mu = \left\{ \mathbf{e} \in \mathbb{R}^n \left| \begin{array}{c} \mathbf{ee}^T \preccurlyeq E[\mathbf{e}_k^a]E[\mathbf{e}_k^a]^T + \gamma Cov(\mathbf{e}_k^a), \\ \mathbf{e}_k^a = \mathbf{e}_k^a(\mathbf{a}_{1..k}), \ \mathbf{a}_{1..k} \in \mathcal{A}_k^\mu \end{array} \right. \right\}. \quad (42)$$

Furthermore, the global policy-aware reachable region $\mathcal{R}^\mu$ of the state estimation error $\mathbf{e}_k^a$ is the set

$$\mathcal{R}^\mu = \cup_{k=0}^\infty \mathcal{R}_k^\mu. \quad (43)$$

The above definition introduces a region that can be reached by an attacker that both considers past behavior and plans accordingly into the future to avoid being detected. Since $\mathcal{A}_k^\mu \subseteq \mathcal{A}_k$, it directly follows that $\mathcal{R}^\mu \subseteq \mathcal{R}$, and the boundedness property holds. Finally, note that when no integrity enforcements are used it follows that $\mathcal{R}^\mu \equiv \mathcal{R}$.

1) Guarantees with Sensor-wise Integrity Enforcement: Due to space constraint, we now consider the case where the system has one unstable eigenvalue λ_1 with the corresponding eigenvector $\mathbf{v_1}$, but the result can be generalized. Also, let's assume that all sensor integrity enforcement policies use $f_i = 1$ and have $t_k^i = t_k^{i+1} - 1$ for all k and all $i = 1, ..., p - 1$

(i.e., sensors enforce integrity in consecutive points, first s_1, then s_2, etc); this also implies all L_i are equal.

It can be shown that the system is not perfectly attackable in this case. The proof follows the ideas from the proofs of Theorems 4 and 5. If s_1 integrity is enforced at $t_k^1 = j$, that would mean that $\Delta z_{j,1} = P_{\{s_1\}} \mathbf{CA} \Delta \mathbf{e}_{j-1} = P_{\{s_1\}} \tilde{\mathbf{C}} \mathbf{J} \alpha_{1..n}$ as in Theorem 5, and thus $\|P_{\{s_1\}} \tilde{\mathbf{C}} \mathbf{J} \alpha_{1..n}\|_2 \leq 1$. From Lemma 3, if $\Delta \mathbf{e}_{j-1}$ is unbounded, only $\alpha_1 \to \infty$, and thus $\mathbf{J}_1$ as in (38) is scalar. To account for this, $P_{\{s_1\}} \mathbf{C} \mathbf{v}_1 \lambda_1 \alpha_1$ has to be zero, which implies that $\mathbf{v}_1 \in null(P_{\{s_1\}} \mathbf{C})$. Similarly, it can be shown that from $\Delta \mathbf{z}_{j,i}$, it follows that $\mathbf{v}_1 \in null(P_{\{s_i\}} \mathbf{C})$ for $1 \leq i \leq n$. This can be represented as $\mathbf{C} \mathbf{v}_1 = \mathbf{0}$, which since $\lambda_1 \neq 0$ implies that $\mathbf{v}_1 \in null(\mathbf{C})$. This is a contradiction because $(\mathbf{A}, \mathbf{C})$ is observable from our initial assumptions.

V. ANALYSIS AND DESIGN OF SAFE INTEGRITY ENFORCEMENT POLICIES

The previous section shows that with even intermittent integrity enforcements, a stealthy attacker cannot introduce an unbounded state estimation error, irrelevant of the set of compromised sensors $\mathcal{K}$. However, we still need to provide a method to evaluate whether a specific integrity enforcement policy ensures the desired estimation performance (i.e., state estimation error) even in the presence of attacks. Our goal is to also provide a framework to derive integrity enforcement policies ensuring that the estimation errors remain within a desired region even under attack. Thus, in this section, we introduce a computationally efficient method to achieve this based on an efficient estimation of the reachable region $\mathcal{R}_k^\mu$ from (42) for systems with intermittent data integrity enforcements.

A. Estimation Errors with Intermittent Integrity Enforcements

Consider an LTI system from (1), (9) with a global data integrity policy (μ, f, L). As in Definition 5, we use $\mathbf{a}_{1..k} = [(\mathbf{a}_1)^T \ ... \ (\mathbf{a}_k)^T]^T \in \mathbb{R}^{pk}$ to capture attack vectors up to step k, where $supp(\mathbf{a}_j) = \tilde{\mathcal{K}}_j$, $j = 1, ..., k$, and

$$\tilde{\mathcal{K}}_j = \begin{cases} \emptyset, & j - i \in \mu, \text{ for some } i, \ 0 \leq i < f, \\ \mathcal{K}, & \text{otherwise} \end{cases}$$

Here, $\tilde{\mathcal{K}}_j$ captures the set of compromised sensor measurements received in step j – i.e., if data integrity is enforced at step j then no measurements are compromised. Let us define $supp(\mathbf{a}_{1..k}) = \mathcal{Q}_k \subseteq \{1, ..., pk\}$; note that $\mathcal{Q}_k$ captures information about the applied integrity enforcement policy, and

$$|\mathcal{Q}_k| = |\mathcal{Q}_{k-1}| + |\tilde{\mathcal{K}}_k| = \sum_{i=1}^{k} |\tilde{\mathcal{K}}_i|. \tag{44}$$

From (10) and (11), $\Delta \mathbf{e}_k$ and $\Delta \mathbf{z}_k$ can be captured as

$$\Delta \mathbf{e}_k = -\underbrace{\left[(\mathbf{A} - \mathbf{KCA})^{k-1} \mathbf{K} \mid ... \mid \mathbf{K} \right]}_{\mathbf{M}_k} \mathbf{a}_{1..k}$$

$$\Delta \mathbf{z}_k = \underbrace{\left[-\mathbf{CAM}_{k-1} \mid \mathbf{I} \right]}_{\mathbf{N}_k} \mathbf{a}_{1..k} \tag{45}$$

To incorporate the information about the sparsity of the attack vector, we use suitable projections onto $\mathcal{Q}_k$ and $\tilde{\mathcal{K}}_1, ..., \tilde{\mathcal{K}}_k$, which satisfy $\mathbf{P}_{\mathcal{Q}_k} = BlckDiag(\mathbf{P}_{\tilde{\mathcal{K}}_1}, ..., \mathbf{P}_{\tilde{\mathcal{K}}_k})$. In addition, it holds that $\mathbf{P}_{\tilde{\mathcal{K}}_j}^\dagger = \mathbf{P}_{\tilde{\mathcal{K}}_j}^T$, since $\mathbf{P}_{\tilde{\mathcal{K}}_j} \mathbf{P}_{\tilde{\mathcal{K}}_j}^T = I_{|\tilde{\mathcal{K}}_j|}$, for $j = 1, ..., k$, and thus $\mathbf{P}_{\mathcal{Q}_k}^\dagger = \mathbf{P}_{\mathcal{Q}_k}^T$. Then, (45) can be restated as

$$\Delta \mathbf{e}_k = -\underbrace{\left[(\mathbf{A} - \mathbf{KCA})^{k-1} \mathbf{KP}_{\tilde{\mathcal{K}}_1}^\dagger \mid ... \mid \mathbf{KP}_{\tilde{\mathcal{K}}_k}^\dagger \right]}_{\mathbf{M}_k \mathbf{P}_{\mathcal{Q}_k}^\dagger} \mathbf{P}_{\mathcal{Q}_k} \mathbf{a}_{1..k}$$

$$\tag{46}$$

$$\Delta \mathbf{z}_k = \underbrace{\left[-\mathbf{CAM}_{k-1} \mathbf{P}_{\mathcal{Q}_{k-1}^\dagger} \mid \mathbf{P}_{\tilde{\mathcal{K}}_k}^\dagger \right]}_{\mathbf{N}_k \mathbf{P}_{\mathcal{Q}_k}^\dagger} \mathbf{P}_{\mathcal{Q}_k} \mathbf{a}_{1..k}$$

with matrices $\mathbf{M}_k \mathbf{P}_{\mathcal{Q}_k^\dagger}$ and $\mathbf{N}_k \mathbf{P}_{\mathcal{Q}_k^\dagger}$ capturing information about the time steps in which data integrity is enforced.

For the general form of the detection function g_k, it may not be possible to obtain a simple analytical solution for the regions $\mathcal{R}_k^\mu$ and $\mathcal{R}^\mu$. Therefore, in this section we will focus on a specific log-likelihood ratio test as used in [10], [12]; however, the presented method can also be extended in a similar fashion to cover other detectors such as generalized likelihood tests. Specifically, we consider the detection function captured as

$$g_k = g_{k-1} + \Lambda_k = \sum_{\tau=1}^{k} (\frac{1}{2} \mathbf{z}_\tau^T \mathbf{Q}^{-1} \mathbf{z}_\tau + log\ c\sqrt{(2\pi)^p det(\mathbf{Q})}) =$$

$$= \frac{1}{2} \sum_{\tau=1}^{k} (\mathbf{z}_\tau^T \mathbf{Q}^{-1} \mathbf{z}_\tau) + k\ log\ c\sqrt{(2\pi)^p det(\mathbf{Q})}.$$

$$\tag{47}$$

Here, $\Lambda_k = log \frac{f_a(\mathbf{z}_k)}{f(\mathbf{z}_k)}$ with f_a and f being probability density functions of the residuals under the attack and in regular operation respectively, and $c = e^{-\frac{p}{2}} / \sqrt{(2\pi)^p det(\mathbf{Q})}$ is a design constant initialized such that log-likelihood ratio $\Lambda_k \equiv 0$ when the system is not under attack [12]. Therefore, in this case the attacker's stealthiness constraint from (16) (i.e., $P(g_k^a > h) \leq P(g_k > h) + \varepsilon$) can be captured as

$$P\left(\sum_{\tau=1}^{k}(\mathbf{z}_\tau^{aT} \mathbf{Q}^{-1} \mathbf{z}_\tau^a) > 2h + kp\right) \leq \varepsilon + P\left(\sum_{\tau=1}^{k}(\mathbf{z}_\tau^T \mathbf{Q}^{-1} \mathbf{z}_\tau) > 2h + kp\right)$$

Given that these two sums have the non-central χ^2 (left) and (central) χ^2 distributions, from Theorem 1 and the proof of Lemma 1 it follows that the above constraint is equivalent to

$$\sqrt{\sum_{\tau=1}^{k} \|\Delta \mathbf{z}_\tau\|_{\mathbf{Q}^{-1}}^2} \leq \alpha_{\chi^2}(\varepsilon, kp, 2h + kp). \tag{48}$$

On the other hand, from (46) it follows that

$$\sum_{\tau=1}^{k} \|\Delta \mathbf{z}_\tau\|_{\mathbf{Q}^{-1}}^2 = \sum_{\tau=1}^{k} \Delta \mathbf{z}_\tau^T \mathbf{Q}^{-1} \Delta \mathbf{z}_\tau =$$

$$= \sum_{\tau=1}^{k} (\mathbf{P}_{\mathcal{Q}_k} \mathbf{a}_{1..k})^T [\mathbf{N}_\tau \mathbf{P}_{\mathcal{Q}_\tau^\dagger} \quad \mathbf{0}_{p \times (|\mathcal{Q}_k| - |\mathcal{Q}_\tau|)}]^T \mathbf{Q}^{-1}$$

$$[\mathbf{N}_\tau \mathbf{P}_{\mathcal{Q}_\tau^\dagger} \quad \mathbf{0}_{p \times (|\mathcal{Q}_k| - |\mathcal{Q}_\tau|)}] \mathbf{P}_{\mathcal{Q}_k} \mathbf{a}_{1..k}.$$

Hence, from (48), the attacker's stealthiness constraint under considered integrity enforcement policy μ can be captured as

$$\|P_{\mathcal{Q}_k}\mathbf{a}_{1..k}\|_{\Theta_k} \leq \alpha_{\chi^2}(\varepsilon, kp, 2h + kp), \tag{49}$$

where

$$\Theta_k = \sum_{\tau=1}^{k}[\mathbf{N}_\tau \mathbf{P}_{\mathcal{Q}_\tau}^\dagger \ \mathbf{0}_{p\times(|\mathcal{Q}_k|-|\mathcal{Q}_\tau|)}]^T \mathbf{Q}^{-1}[\mathbf{N}_\tau \mathbf{P}_{\mathcal{Q}_\tau}^\dagger \ \mathbf{0}_{p\times(|\mathcal{Q}_k|-|\mathcal{Q}_\tau|)}]. \tag{50}$$

For the above matrix Θ_k, the following property holds.

Lemma 4: For any $k \geq 1$, the matrix Θ_k is positive definite.

Proof: We start with the case when $k = 1$. From Definition 4, data integrity is not enforced at $k = 1$ and thus $\mathcal{Q}_1 = \hat{\mathcal{K}}_1 = \mathcal{K}$. Due to the way projection matrices are formed, we have that

$$\mathbf{P}_{\mathcal{Q}_1}^{\dagger T}\mathbf{P}_{\mathcal{Q}_1}^\dagger = \mathbf{I}_{|\mathcal{Q}_1|} \succ 0 \quad \text{and} \quad \Theta_1 = [\mathbf{P}_{\mathcal{Q}_1}^\dagger]^T\mathbf{Q}^{-1}[\mathbf{P}_{\mathcal{Q}_1}^\dagger].$$

Since $\mathbf{Q} \succ 0$, it follows that $\Theta_1 \succ 0$ as well.

Now, consider the case $k \geq 2$ and let us assume that Θ_{k-1} is positive definite. From (50) it follows that

$$\Theta_k = \underbrace{\begin{bmatrix} \Theta_{k-1} & \mathbf{0}_{|\mathcal{Q}_{k-1}|\times|\tilde{\mathcal{K}}_k|} \\ \mathbf{0}_{|\tilde{\mathcal{K}}_k|\times|\mathcal{Q}_{k-1}|} & \mathbf{0}_{|\tilde{\mathcal{K}}_k|\times|\tilde{\mathcal{K}}_k|} \end{bmatrix}}_{\tilde{\Theta}_{k-1}} + \underbrace{[\mathbf{N}_k P_{\mathcal{Q}_k}^\dagger]^T\mathbf{Q}^{-1}[\mathbf{N}_k P_{\mathcal{Q}_k}^\dagger]}_{\tilde{\Theta}_k} \tag{51}$$

and we consider the following two cases.

Case I: There does not exist i, such that $0 \leq i < f$ and $k - i \in \mu$; this implies that integrity is not enforced at the step k and $\hat{\mathcal{K}}_k = \mathcal{K}$. Because both $\Theta_{k-1} \succ 0$ and $\mathbf{Q} \succ 0$, both addends in (51) are positive semidefinite matrices, and $\Theta_k \succeq 0$. In addition, since Θ_{k-1} is positive definite by assumption, $null\left(\tilde{\Theta}_{k-1}\right) = \mathcal{R}\left([\mathbf{0}_{|\tilde{\mathcal{K}}_k|\times|\mathcal{Q}_{k-1}|} \ \mathbf{I}_{|\tilde{\mathcal{K}}_k|}]^T\right)$. Furthermore, from (46), we have (52). Given that $(\mathbf{P}_{\tilde{\mathcal{K}}_k}^\dagger)^T\mathbf{Q}^{-1}\mathbf{P}_{\tilde{\mathcal{K}}_k}^\dagger \succ 0$, it follows that $null\left(\tilde{\Theta}_k\right)$ cannot have non-zero vectors from $\mathcal{R}\left([\mathbf{0}_{|\tilde{\mathcal{K}}_k|\times|\mathcal{Q}_{k-1}|} \ \mathbf{I}_{|\tilde{\mathcal{K}}_k|}]^T\right)$. Therefore,

$$null(\tilde{\Theta}_k) \cap null(\tilde{\Theta}_{k-1}) = \{\mathbf{0}\}. \tag{53}$$

Now, assume that there exists a non-zero vector $\mathbf{v}$ such that $\mathbf{v} \in null(\Theta_k)$ – i.e., $(\tilde{\Theta}_k + \tilde{\Theta}_{k-1})\mathbf{v} = 0$, and thus

$$\mathbf{v}^T\tilde{\Theta}_k\mathbf{v} = -\mathbf{v}^T\tilde{\Theta}_{k-1}\mathbf{v}.$$

However, since $\mathbf{v}$ cannot be in the null-spaces of both matrices due to (53), and $\tilde{\Theta}_{k-1}$ and $\tilde{\Theta}_k$ are both positive semidefinite, this is a clear contradiction. Consequently, $\Theta_k = \tilde{\Theta}_k + \tilde{\Theta}_{k-1}$ $null(\Theta_k) = \{\mathbf{0}\}$, and since Θ_k is a positive semidefinite matrix it holds that $\Theta_k \succ 0$.

Case II: There exists i, such that $0 \leq i < f$ and $k - i \in \mu$; i.e., integrity is enforced at the step k. Thus, $|\hat{\mathcal{K}}_k| = 0$, so $\tilde{\Theta}_{k-1} = \Theta_{k-1}$ is positive definite. Hence, since $\tilde{\Theta}_k \succeq 0$, it follows that $\Theta_k = \tilde{\Theta}_k + \tilde{\Theta}_{k-1}$ is positive definite. ∎

Now, the specification of the stealthiness condition from (49) allows us to obtain the following result.

Theorem 7: The k-reachable region $\mathcal{R}_k^\mu$ under a global data integrity enforcement policy (μ, f, L) can be represented as

$$\mathcal{R}_k^\mu = \left\{\mathbf{e}_k^a | \mathbf{e}_k^a\mathbf{e}_k^{aT} \preccurlyeq \alpha_{\chi^2}^2[\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0}]\Theta_t^{-1}[\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0}]^T + \gamma\Sigma\right\} \tag{54}$$

where $\alpha_{\chi^2}^2 = \alpha_{\chi^2}^2(\varepsilon, tp, 2h + tp)$, t is the first end of an integrity enforcement block following k – i.e., the earliest time point such that $t - f + 1 \in \mu$ and $k \leq t$, and $\mathbf{0} = \mathbf{0}_{|\mathcal{Q}_k|\times(|\mathcal{Q}_t|-|\mathcal{Q}_k|)}$.

Proof: Consider the stealthiness constraints (49) at time t, which can be written as

$$\alpha_{\chi^2}^2(\varepsilon, tp, 2h + tp) - (\mathbf{P}_{\mathcal{Q}_t}\mathbf{a}_{1..t})^T\Theta_t\mathbf{P}_{\mathcal{Q}_t}\mathbf{a}_{1..t} \geq 0. \tag{55}$$

Now, using Schur complement and Lemma 4, we obtain

$$\begin{bmatrix} \Theta_t^{-1} & \mathbf{P}_{\mathcal{Q}_t}\mathbf{a}_{1..t} \\ (\mathbf{P}_{\mathcal{Q}_t}\mathbf{a}_{1..t})^T & \alpha_{\chi^2}^2(\varepsilon, tp, 2h + tp) \end{bmatrix} \succcurlyeq 0 \tag{56}$$

As the left hand side of (56) is positive semidefinite, when multiplied by a matrix from the left, and its transpose from the right, this product will also be positive semidefinite. If we use the projection matrix $\mathbf{P}_{\{1,...,k,t+1\}}$ for this, we effectively reduce the matrix from (56) by removing pairs of rows and columns corresponding to $\mathbf{a}_{k+1..t}$. Thus, we obtain that

$$\begin{bmatrix} [\mathbf{I}_{|\mathcal{Q}_k|} \ \mathbf{0}]\Theta_t^{-1}[\mathbf{I}_{|\mathcal{Q}_k|} \ \mathbf{0}]^T & \mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k} \\ (\mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k})^T & \alpha_{\chi^2}^2(\varepsilon, tp, 2h + tp) \end{bmatrix} \succcurlyeq 0 \tag{57}$$

where $\mathbf{0} = \mathbf{0}_{|\mathcal{Q}_k|\times(|\mathcal{Q}_t|-|\mathcal{Q}_k|)}$. Furthermore, with condition (57) we need to compute only single Θ_t^{-1} for all points between integrity enforcement blocks, as constraints for prior attacks (i.e., time points before t) directly follow from (57).

The LMI in (58) follows from (57) as it forms a quadratic representation. We use this specific matrix as it allow us to argue about the stealthiness condition using $\Delta\mathbf{e}_k$ rather than $\mathbf{a}_{1..k}$. Using (46) and Schur complement once again, we have

$$[\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0}]\Theta_t^{-1}[(\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0})]^T - \frac{1}{\alpha_{\chi^2}^2}\Delta\mathbf{e}_k\Delta\mathbf{e}_k^T \succcurlyeq 0, \tag{59}$$

where $\alpha_{\chi^2}^2 = \alpha_{\chi^2}^2(\varepsilon, tp, 2h + tp)$. Hence, from (59) and the definition of $\mathcal{R}_k^\mu$ from (42), as well as (12) and the fact that $Cov[\mathbf{e}_k^a] = \Sigma$, we finally obtain that (54) holds. ∎

The representation of the reachable set from (54) can be simplified further. Let's define $\mathbf{Y}_k$ as

$$\mathbf{Y}_k = \alpha_{\chi^2}^2(\varepsilon, tp, 2h+tp)[\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0}]\Theta_t^{-1}[\mathbf{M}_k P_{\mathcal{Q}_k}^\dagger \ \mathbf{0}]^T + \gamma\Sigma. \tag{60}$$

Then (54) is equivalent to $\mathbf{Y}_k - \mathbf{e}_k^a\mathbf{e}_k^{aT} \succcurlyeq 0$, and thus by using Schur complement we obtain an alternative representation of the k-reachable regions as

$$\mathcal{R}_k^\mu = \{\mathbf{e}_k^a | \mathbf{e}_k^{aT}\mathbf{Y}_k^{-1}\mathbf{e}_k^a \preccurlyeq 1\}, \tag{61}$$

for the positive definite matrix $\mathbf{Y}_k$ defined in (60). The above representation can be exploited for efficient computation of the reachable-regions.

Furthermore, as we described in Section II-B, the attacker's goal is to maximize the expected state estimation error

$$\tilde{\boldsymbol{\Theta}}_k = \begin{bmatrix} (\mathbf{CAM}_{k-1}\mathbf{P}^\dagger_{\mathcal{Q}_{k-1}})^T\mathbf{Q}^{-1}\mathbf{CAM}_{k-1}\mathbf{P}^\dagger_{\mathcal{Q}_{k-1}} & (\mathbf{CAM}_{k-1}\mathbf{P}^\dagger_{\mathcal{Q}_{k-1}})^T\mathbf{Q}^{-1}\mathbf{P}^\dagger_{\tilde{\mathcal{K}}_k} \\ (\mathbf{P}^\dagger_{\tilde{\mathcal{K}}_k})^T\mathbf{Q}^{-1}\mathbf{CAM}_{k-1}\mathbf{P}^\dagger_{\mathcal{Q}_{k-1}} & (\mathbf{P}^\dagger_{\tilde{\mathcal{K}}_k})^T\mathbf{Q}^{-1}\mathbf{P}^\dagger_{\tilde{\mathcal{K}}_k} \end{bmatrix} \tag{52}$$

$$\begin{bmatrix} -\mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k} & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times|\mathcal{Q}_k|} & 1 \end{bmatrix}\begin{bmatrix} [\mathbf{I}\;\;\mathbf{0}]\boldsymbol{\Theta}_t^{-1}[\mathbf{I}\;\;\mathbf{0}]^T & \mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k} \\ (\mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k})^T & \alpha^2_{\chi^2}(\varepsilon,tp,2h+tp) \end{bmatrix}\begin{bmatrix} -\mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k} & \mathbf{0}_{n\times 1} \\ \mathbf{0}_{1\times|\mathcal{Q}_k|} & 1 \end{bmatrix}^T \succcurlyeq 0 \Longleftrightarrow$$

$$\Longleftrightarrow \begin{bmatrix} \mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k}[\mathbf{I}\;\;\mathbf{0}]\boldsymbol{\Theta}_t^{-1}[\mathbf{I}\;\;\mathbf{0}]^T(\mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k})^T & -\mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k}\mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k} \\ -(\mathbf{P}_{\mathcal{Q}_k}\mathbf{a}_{1..k})^T(\mathbf{M}_k\mathbf{P}^\dagger_{\mathcal{Q}_k})^T & \alpha^2_{\chi^2}(\varepsilon,tp,2h+tp) \end{bmatrix} \succcurlyeq 0. \tag{58}$$

$E[\mathbf{e}_k^a] = \Delta\mathbf{e}_k$. From the above discussion, the following corollary directly holds by considering the case when $\gamma = 0$.

Corollary 4: At any time k, the maximal norm of the expected state estimation error $\mathbf{e}_k^a$ caused by the attack satisfies

$$\max\|E[\mathbf{e}_k^a]\|_2 = \frac{1}{\sqrt{\lambda_{max}(\tilde{\mathbf{Y}}_k)}}, \tag{62}$$

where $\lambda_{max}(\tilde{\mathbf{Y}}_k)$ denotes the largest eigenvalue of the matrix $\tilde{\mathbf{Y}}_k = \alpha^2_{\chi^2}(\varepsilon,tp,2h+tp)[\mathbf{M}_k P^\dagger_{\mathcal{Q}_k}\;\;\mathbf{0}]\boldsymbol{\Theta}_t^{-1}[\mathbf{M}_k P^\dagger_{\mathcal{Q}_k}\;\;\mathbf{0}]^T$, and t is the next end of integrity enforcement block – i.e., the earliest time point such that $t - f + 1 \in \mu$ and $k \leq t$.

The above corollary provides a very efficient way to evaluate worst-case effects of attacks when an intermittent data integrity enforcement policy is used. By quantifying degradation of the expected state estimation error in the presence of attacks we can analyze the impact of the integrity enforcement policy on limiting the attacker, which can then be used for design of suitable integrity enforcement policies.

B. Design of Periodic Integrity Enforcement Policies

For policy design, it is necessary to be able to evaluate impact of an integrity enforcement policy μ, not only on reachable regions $\mathcal{R}_k^\mu$, for any k, but even more importantly on $\mathcal{R}^\mu$ from (43). To achieve this, we have to obtain the terminating value t from Theorem 7, or equivalently from (61), such that the reachability analysis can be completed after $\mathcal{R}_t^\mu$ is obtained – i.e., for which $\mathcal{R}^\mu = \mathcal{R}_{1..t}^\mu$, where $\mathcal{R}_{1..t}^\mu = \bigcup_{k=1}^t \mathcal{R}_k^\mu$. In the general case, the analysis may never terminate, depending on the particular policy (μ, f, L). Therefore, to simplify the analysis, in this section we focus on periodic integrity enforcement policies introduced in Remark 4.

For a periodic integrity enforcement policy (μ, f, L), consider t_1 and $t_2 = t_1 + L$ time points at which consecutive integrity enforcement blocks end – i.e., $t_1 - f + 1 \in \mu$ and $t_2 - f + 1 \in \mu$. From the proof of Theorem 7, if the stealthiness requirements from the condition in (56) are satisfied at any time $t \in \mu$, then they are satisfied for all $k < t$, since (57) follows from (56). Given that $\mathbf{a}_{t_1-f+1} = ... = \mathbf{a}_{t_1} = \mathbf{0}$ and $\mathbf{a}_{t_2-f+1} = ... = \mathbf{a}_{t_2} = \mathbf{0}$, and that the stealthiness requirements remain consistent throughout the analysis, it follows that the evolution of the estimation error between two consecutive integrity enforcement blocks will depend only on $E[\mathbf{e}_{t_1}^a] = \Delta\mathbf{e}_{t_1}$ and $E[\mathbf{e}_{t_2}^a] = \Delta\mathbf{e}_{t_2}$, or more specifically

$\mathcal{R}_{t_1}^\mu$ and $\mathcal{R}_{t_2}^\mu$. Thus, if $\mathcal{R}_{t_2}^\mu \subseteq \mathcal{R}_{t_1}^\mu$ and $\mathcal{R}_{1..t_2}^\mu \subseteq \mathcal{R}_{1..t_1}^\mu$, then no new estimation error values can be reached after time t_2 and the terminating time for the reachability analysis can be t_1, since after time t_2 as well as after all following ends of integrity enforcement blocks the state estimation errors would start from a subset of the error values from $\mathcal{R}_{t_1}^\mu$. In addition, when the above terminating condition is satisfied, the global reachable region of the state estimation error can be obtained as $\mathcal{R}^\mu = \bigcup_{k=1}^\infty \mathcal{R}_k^\mu = \bigcup_{k=1}^{t_1} \mathcal{R}_k^\mu = \mathcal{R}_{1..t_1}^\mu$.

Consequently, using Algorithm 1 we can compute a periodic integrity enforcement policy that maximizes L (i.e., reduces the integrity enforcement rate) while limiting the attacker's influence. Specifically, the algorithm will result in the enforcement policy that ensures that the state of reachable estimation errors does not contain points outside the set of safe (i.e., acceptable) errors $\mathcal{R}_{\mathbf{e}^a}$. In our evaluations in the next section, we define $\mathcal{R}_{\mathbf{e}^a}$ using a threshold $\|\Delta\mathbf{e}_{max}\|_2$ for the maximal 2-norm of the expected state estimation error due to attacks. Thus, the safety condition in Line 18 of the algorithm is mapped into $max(\|\mathbf{e}_1^a\|_{max},...,\|\mathbf{e}_t^a\|_{max}) \geq \|\Delta\mathbf{e}_{max}\|_2$, where $\|\mathbf{e}_k^a\|_{max} = \max\|E[\mathbf{e}_k^a]\|_2$ as computed in (62).

Finally, while we do not provide any guarantees that Algorithm 1 will always terminate, for all analyzed systems, including the case studies from the next section, the condition in Line 17 was always eventually satisfied. Therefore, for all considered systems we have been able to use the algorithm to obtain periodic integrity enforcement policies that ensure desired estimation performance even in the presence of attacks.

VI. CASE STUDIES

In this section, on automotive case studies we illustrate how intermittent data integrity enforcements can ensure satisfiable control performance even in the presence of attacks. For both studies, sensor values are transmitted over an internal vehicle's network, such as commonly used CAN bus. Note that in [14], we provide additional automotive case-studies (and the overall scheduling framework) for intermittent authentication of CAN-bus messages from system sensors, and in [15] we show benefits of intermittent authentication on vehicle's ECU scheduling.

A. Case Study: Vehicle Trajectory Following

We start with the model used in [9] to describe vulnerabilities and potential attacks on autonomous systems adapted for

Algorithm 1 Procedure for design of periodic integrity enforcement policies.

Inputs: System model, safe reachable region $\mathcal{R}_{\mathbf{e}^a}$ for the state estimation error $\mathbf{e}^a$

1: Enforcement distance $L = 0$
2: **repeat**
3:     $L = L + 1$
4:     Form policy (μ, f, L) such that distance between consecutive elements in μ is L and $t_0 = L$
5:     Assign $t = 0$ and the reachable region $\mathcal{R}_{1..t} = \varnothing$
6:     **repeat**
7:         $t_{old} = t$
8:         $\mathcal{R}_{1..t_{old}} = \mathcal{R}_{1..t}$
9:         $t = min\{t'|t' \in \mu \wedge t' > t_{old}\}$
10:        Compute $\mathbf{N}_{t_{old}+1}, ..., \mathbf{N}_t, \mathbf{M}_{t_{old}+1}, ..., \mathbf{M}_t$ from (45)
11:        Compute $\mathbf{\Theta}_t$ from (50)
12:        Compute $\alpha(\varepsilon, tp, 2h + tp)$
13:        **for** $k = t_{old} + 1, \ldots, t$ **do**
14:            Compute $\mathcal{R}_k^\mu$ using (61)
15:            $\mathcal{R}_{1..t} = \mathcal{R}_{1..t} \cup \mathcal{R}_k^\mu$
16:        **end for**
17:     **until** $\mathcal{R}_{1..t} \subseteq \mathcal{R}_{1..t_{old}}$ and $\mathcal{R}_t \subseteq \mathcal{R}_{t_{old}}$
18: **until** $\mathcal{R}_{1..t_{old}} \setminus \mathcal{R}_{\mathbf{e}^a} \neq \varnothing$
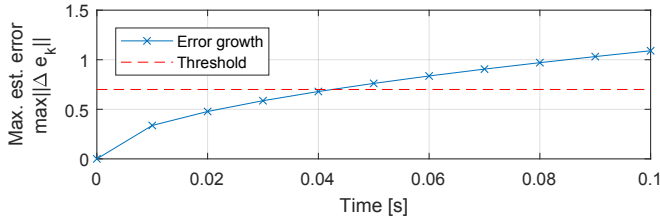19: Accept policy $(\mu, f, L - 1)$

Fig. 4. Evolution of the maximal estimation error for vehicle tracking; without integrity enforcements, the attacker forces the system outside of the safe range in 4 steps.

two-axis tracking; we obtain the following discretized models (with sampling period of $0.01s$) for each axis

$$\mathbf{A}_d = \begin{bmatrix} 1 & 0.01 \\ 0 & 1 \end{bmatrix} \quad \mathbf{B}_d = \begin{bmatrix} 0.0001 \\ 0.01 \end{bmatrix} \quad \mathbf{C}_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (63)$$

Assume that the attacker can modify the values from all sensors. The system is perfectly attackable as the matrix $\mathbf{A}_d$ is unstable and $supp(\mathbf{C}v) \in \mathcal{K}$, since $\mathcal{K} = \mathcal{S}$.

We consider the largest additive estimation error on position to be $0.5\ m$ and on speed to be $0.5\ \frac{m}{s}$, resulting in $\|\Delta \mathbf{e}_{max}\|_2 = 0.7$. We also set h such that the probability of false positive from (8) is $\beta = 1.5\%$, and additional probability of detection introduced by the attacker from (16) to $\varepsilon = 0.1\%$.

Without integrity enforcements, the attacker could force the state estimation error above $\|\Delta \mathbf{e}_{max}\|_2$ threshold after 4 steps, as shown in Figure 4. We considered three periodic integrity enforcement policies with $f = 1$ as specified in conditions of Theorem 4, and periods $L = 20, 30$ and 35, denoted by μ_{20}, μ_{30}, and μ_{35} respectively. Using results from Section V, we show that the first two policies are safe, while the third policy can violate the $\|\Delta \mathbf{e}_{max}\|_2$ threshold – Figure 5 illustrates the
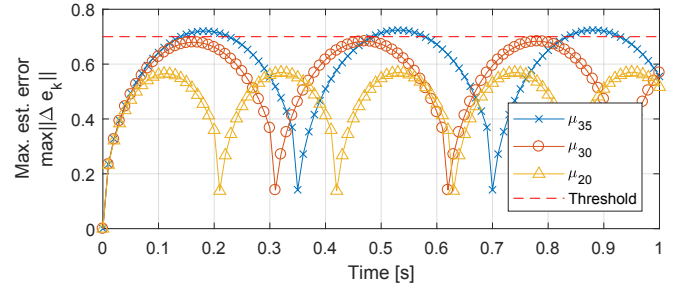

Fig. 5. Maximal estimation error in the presence of attacks on all sensors for vehicle-tracking case study with three different integrity enforcement policies with $f = 1$ and periods $L = 20, 30, 35$.

evolution of the maximal estimation errors for each policy.

Finally, we evaluate the effects of intermittent integrity guarantees for trajectory following on a circular path with $100\ m$ radius, at speed of $3.14\ \frac{m}{s}$. Figure 6 shows results of $200\ s$ long simulations, with attacks starting at $100\ s$. As illustrated, when integrity is enforced on less than 3.4% of messages, i.e. when μ_{30} is employed, we have strong control performance guarantees in the presence of attacks on all vehicle sensors.

### B. Degraded Cooperative Adaptive Cruise Control (dCACC)

Cooperative Adaptive Cruise Control (CACC) employs communication to obtain smaller following distance and better platooning stability than standard Adaptive Cruise Control. To achieve this, each vehicle is equipped with a lidar and acceleration measurement sent from the preceding vehicle. However, when acceleration data is not available CACC needs to switch to dCACC, that is based only on local vehicle measurements. In this mode, Singer acceleration model is used to estimate acceleration of the preceding vehicle [29] – i.e.,

$$\begin{bmatrix} \dot{d} \\ \dot{v} \\ \dot{a} \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{\tau} \end{bmatrix} \begin{bmatrix} d \\ v \\ a \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} u \end{bmatrix} \quad (64)$$

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} d \\ v \\ a \end{bmatrix}. \quad (65)$$

Here, d denotes the distance of the vehicle from the preceding vehicle, v is its speed – both computed from lidar measurements and transmitted over the bus, a is the acceleration, u is the control input, while $\tau = 0.8$ represents maneuver time constant of the preceding vehicle [29]. We focus on the cases when the attacker compromises all car sensors, making the system perfectly attackable. We set maximal estimation error to be $0.5m$ on position, $3.3\frac{m}{s}$ on speed, and $0.3\frac{m}{s^2}$ on acceleration, resulting in $\|\Delta \mathbf{e}_{max}\|_2 = 3.351$.

As in trajectory tracking, we assume $\varepsilon = 0.1\%$, and $\beta = 0.35\%$. Since observability index $\psi = 2$ and number of unstable eigenvalues of $\mathbf{A}$ is 2, then $f = 2$. For periodic policy with $L = 20$ we obtain the maximal reachable estimation errors in the presence of stealthy attacks as presented in Figure 7. In addition, visual representation of reachable regions with this policy in comparison to a system without integrity enforcement
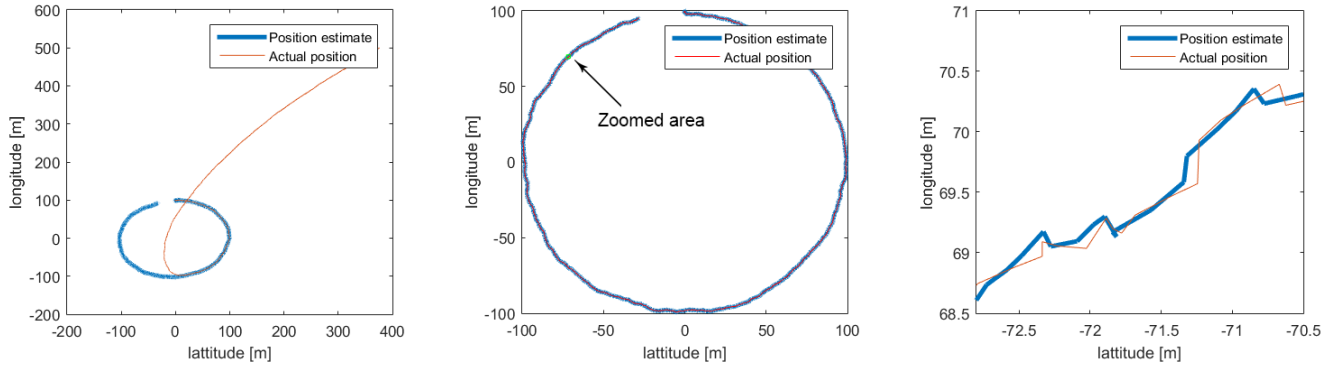
(a) State Estimates for system under stealthy attack, without integrity enforcement policies

(b) State estimates under stealthy attack with integrity enforcement policies.

(c) Zoomed section of the Figure 6(b).

Fig. 6. State estimation of the tracked vehicle trajectory - without integrity enforcements a stealthy attacker can introduce a significant estimation error in a short period of time. However, even with intermittent integrity enforcement, the attack effects are negligible. Duration of the simulation is $200\ s$ and the attack starts at $100\ s$.
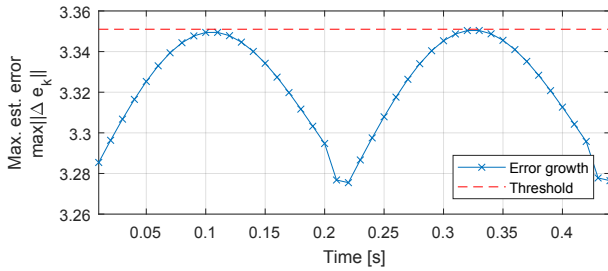

Fig. 7. Evolution of maximal estimation error for dCACC. If we can enforce integrity on two sensor values after every twenty unsecured sensor values, the system remains under the specified safety threshold.
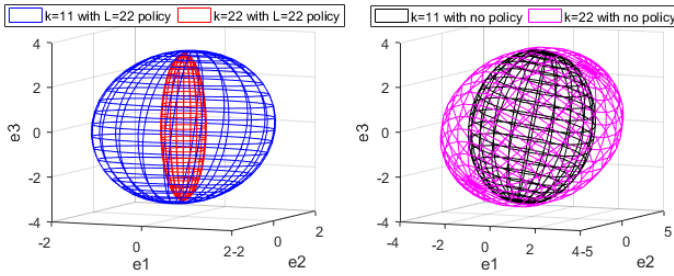

Fig. 8. Reachable state estimation errors in the presence of stealthy attacks for dCACC in steps $k = 11$ and $k = 22$ with and without data integrity enforcement. Without integrity enforcement, the size of reachable regions keeps increasing, while when integrity is being enforced with policy $L = 20$ and $f = 2$, estimation error evolves as in Figure 7, and the attacker is contained between red and blue ellipsoids.

is shown in Figure 8. These results illustrate that even with 10% authenticated messages the system ensures satisfiable performance under false-date injection attacks.

VII. Conclusion

In this paper, we have focused on the problem of network-based attacks on standard linear state estimators. We have considered systems with Kalman filter-based estimators and a general type of residual-based intrusion detectors, covering widely used detectors such as χ^2 detectors. For these systems,

we have studied effects of intermittent data integrity enforcements, such as the use of message authentication codes, on control performance in the presence of attacks. We have shown that when integrity of sensor measurements is enforced only intermittently, a stealthy attacker cannot insert an unbounded state estimation error. In addition, we have introduced a framework that facilitates both evaluation and design of these intermittent policies by providing analysis of the reachable state estimation errors in the presence of stealthy attacks. Although the framework has been developed for systems that employ windowed and log-likelihood ratio detectors, the presented techniques can be extended for detectors from the general class described in Section II. Finally, on three automotive case studies, we have highlighted how devastating stealthy false-data injection attacks can be, and how with the use of intermittent integrity enforcement we can ensure desired control performance with a significant reduction in the communication and computation overhead.

The presented method to analyze the effects of intermittent use of authentication can also provide the foundation for optimal resource allocation in systems where several control loops share communication and computation resources. Although we present some initial results in [14] for bandwidth allocation over a shared network, a more systematic approach to optimal resource allocation with strong Quality-of-Control guarantees in the presence of attacks is an avenue for future work.

References

[1] M. L. Boas. *Mathematical methods in the physical sciences*. Wiley, 2006.

[2] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno, et al. Comprehensive experimental analyses of automotive attack surfaces. In *Proceedings of USENIX Security*, 2011.

[3] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[4] G. H. Golub and C. F. Van Loan. *Matrix computations*. JHU Press, 3rd edition, 1996.

[5] R. Ivanov, M. Pajic, and I. Lee. Attack-resilient sensor fusion for safety-critical cyber-physical systems. *ACM Transactions on Embedded Computing Systems*, 15(1):21:1–21:24, Feb. 2016.

[6] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*. Wiley & Sons, 1995.

[7] I. Jovanov and M. Pajic. Sporadic data integrity for secure state estimation. In *56th IEEE Conference on Decision and Control (CDC)*, pages 163–169, 2017.

[8] J.-N. Juang. *Applied System Identification*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1994.

[9] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys. Unmanned aircraft capture and control via gps spoofing. *Journal of Field Robotics*, 31(4):617–636, 2014.

[10] C. Kwon and I. Hwang. Reachability analysis for safety assurance of cyber-physical systems against cyber attacks. *IEEE Transactions on Automatic Control*, PP(99):1–1, 2017.

[11] C. Kwon, W. Liu, and I. Hwang. Analysis and design of stealthy cyber attacks on unmanned aerial systems. *Journal of Aerospace Information Systems*, 1(8), 2014.

[12] C. Kwon, S. Yantek, and I. Hwang. Real-time safety assessment of unmanned aircraft systems against stealthy cyber attacks. *Journal of Aerospace Information Systems*, pages 1–19, 2015.

[13] R. Langner. Stuxnet: Dissecting a cyberwarfare weapon. *Security & Privacy, IEEE*, 9(3):49–51, 2011.

[14] V. Lesi, I. Jovanov, and M. Pajic. Network scheduling for secure cyber-physical systems. In *2017 IEEE Real-Time Systems Symposium (RTSS)*, pages 45–55, Dec 2017.

[15] V. Lesi, I. Jovanov, and M. Pajic. Security-aware scheduling of embedded control tasks. *ACM Trans. Embed. Comput. Syst.*, 16(5s):188:1–188:21, Sept. 2017.

[16] C.-W. Lin, B. Zheng, Q. Zhu, and A. Sangiovanni-Vincentelli. Security-aware design methodology and optimization for automotive systems. *ACM Trans. on Des. Autom. of Elec. Syst.*, 21(1):18, 2015.

[17] Y. Z. Lun, A. D'Innocenzo, I. Malavolta, and M. D. D. Benedetto. Cyber-physical systems security: a systematic mapping study. *CoRR*, abs/1605.09641, 2016.

[18] F. Miao, M. Pajic, and G. Pappas. Stochastic game approach for replay attack detection. In *IEEE 52nd Annual Conference on Decision and Control (CDC)*, pages 1854–1859, Dec 2013.

[19] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. *IEEE Trans. on Control of Network Systems*, 4(1):106–117, March 2017.

[20] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *49th IEEE Conf. on Decision and Control (CDC)*, pages 5967–5972, 2010.

[21] Y. Mo, T.-H. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. Cyber–physical security of a smart grid infrastructure. *Proceedings of the IEEE*, 100(1):195–209, 2012.

[22] Y. Mo and B. Sinopoli. False data injection attacks in control systems. In *First Workshop on Secure Control Systems, CPS Week*, 2010.

[23] Y. Mo and B. Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61(9):2618–2624, 2016.

[24] Y. Mo, S. Weerakkody, and B. Sinopoli. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *Control Systems*, 35(1):93–109, 2015.

[25] M. Pajic, I. Lee, and G. J. Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1):82–92, March 2017.

[26] M. Pajic, J. Weimer, N. Bezzo, O. Sokolsky, G. J. Pappas, and I. Lee. Design and implementation of attack-resilient cyberphysical systems: With a focus on attack-resilient state estimators. *IEEE Control Systems*, 37(2):66–81, April 2017.

[27] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas. Robustness of attack-resilient state estimators. In *ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, pages 163–174, April 2014.

[28] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[29] J. Ploeg, E. Semsar-Kazerooni, G. Lijster, N. van de Wouw, and H. Nijmeijer. Graceful degradation of cooperative adaptive cruise control. *IEEE Trans. on Int. Trans. Sys.*, 16(1):488–497, 2015.

[30] T. R, C. Murguia, and J. Ruths. Tuning windowed chi-squared detectors for sensor attacks. *CoRR*, abs/1710.02573, 2017.

[31] D. Shepard, J. Bhatti, and T. Humphreys. Drone hack. *GPS World*, 23(8):30–33, 2012.

[32] Y. Shoukry, M. Chong, M. Wakaiki, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, J. P. Hespanha, and P. Tabuada. Smt-based observer design for cyber-physical systems under sensor attacks. In *Int. Conference on Cyber-Physical Systems (ICCPS)*, pages 1–10, 2016.

[33] Y. Shoukry and P. Tabuada. Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, 61(8):2079–2091, Aug 2016.

[34] R. Smith. A decoupled feedback structure for covertly appropriating networked control systems. *Proceedings of IFAC World Congress*, pages 90–95, 2011.

[35] R. S. Smith. Covert misappropriation of networked control systems: Presenting a feedback structure. *Control Systems*, 35(1):82–92, 2015.

[36] S. Sundaram. *Fault-tolerant and secure control systems*. University of Waterloo, Lecture Notes, 2012.

[37] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. Pappas. The Wireless Control Network: Monitoring for Malicious Behavior. In *49th IEEE Conference on Decision and Control (CDC)*, pages 5979–5984, Dec 2010.

[38] A. Teixeira, D. Pérez, H. Sandberg, and K. H. Johansson. Attack models and scenarios for networked control systems. In *Int. Conf on High Confidence Networked Systems (HiCoNS)*, pages 55–64, 2012.

[39] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems*, 35(1):24–45, 2015.

[40] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems*, 35(1):24–45, Feb 2015.

[41] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Capkun. On the requirements for successful gps spoofing attacks. In *18th ACM Conf. on Computer and Com. Security*, CCS, pages 75–86, 2011.

[42] K. Zetter. Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid, Wired Magazine, March 2016.

APPENDIX
PROOF OF LEMMA 3

Proof: From (10) and (11), the system Ξ can be described as

$$\Delta \mathbf{e}_k = \mathbf{A}\Delta \mathbf{e}_{k-1} - \mathbf{K}\Delta \mathbf{z}_k \qquad (66)$$

Thus, due to the stealthiness constraint (30), from perspective of estimation error $\Delta \mathbf{e}_k$ the system Ξ is effectively an unstable system with bounded input $\Delta \mathbf{z}_k$. To show that when the estimation error becomes unbounded, the unbounded parts of the vector would belong to vector subspaces corresponding to unstable modes of $\mathbf{A}$ we start by capturing $\Delta \mathbf{e}_k$ in a non-recursive form as

$$\Delta \mathbf{e}_k = -\sum_{i=0}^{k-1} \mathbf{A}^i \mathbf{K}\Delta \mathbf{z}_{k-i}, \qquad (67)$$

since $\Delta \mathbf{e}_0 = \mathbf{0}$. Also, since eigenvectors and generalized eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{A}$ span $\mathbb{R}^n$, we can decompose the estimation error as

$$\Delta \mathbf{e}_k = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i, \qquad \alpha_i \in \mathbb{R}, i = 1, \ldots, n. \qquad (68)$$

Decomposing $\mathbf{K}\Delta \mathbf{z}_{k-i}$ with the same base vectors $\mathbf{v}_1, \ldots \mathbf{v}_n \in \mathbb{R}^n$, we obtain that

$$-\mathbf{K}\Delta \mathbf{z}_{k-i} = \phi_{k-i,1}\mathbf{v}_1 + \cdots + \phi_{k-i,n}\mathbf{v}_n, \qquad \phi_{k-i,j} \in \mathbb{R}, \quad (69)$$

where for $i = 0, \ldots k-1$ and $j = 1, \ldots n$, due to (30) it holds that $\phi_{k-i,j} \in \mathbb{R}$ are bounded – specifically, for $\mathbf{V} = [\mathbf{v}_1 \ldots \mathbf{v}_n] \in \mathbb{R}^{n \times n}$

$$|\phi_{k-i,j}| \le \phi_{max,i} = \max_{\|\mathbf{z}\|_2 = 1} \mathbf{i}_i^T \mathbf{V}^{-1}(\mathbf{Kz}),$$

where $\mathbf{V}$ is invertible as $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, and $\mathbf{i}_i \in \mathbb{R}^n$ is the projection vector with 1 in i^{th} position and zeros otherwise.

Thus, from (67), (68), and (69), we have

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{j=0}^{k-1} \sum_{i=1}^{n} \phi_{k-j,i}\mathbf{A}^{j-1}\mathbf{v}_i. \qquad (70)$$

We now consider two cases, although Case II is more general (and captures Case I as well), its notation is quite cumbersome.

Case I – When $\mathbf{A}$ is diagonalizable, $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ holds, and thus

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{j=0}^{k-1} \sum_{i=1}^{n} \phi_{k-j,i} \lambda_i^{j-1} \mathbf{v}_i. \tag{71}$$

Since $\mathbf{v}_i$ are linearly independent, we obtain

$$|\alpha_i| = |\sum_{j=0}^{k-1} \phi_{k-j,i} \lambda^{j-1}| \le \phi_{max,i} \sum_{j=0}^{k-1} |\lambda_i|^{j-1}, \tag{72}$$

The right side of (72) will converge when $k \to \infty$ if and only if $|\lambda_i| < 1$, which implies that α_i can have arbitrarily large values only if associated with an eigenvector corresponding to an unstable eigenvalues, while α_i associated with stable eigenvalues is bounded.

Case II – In general $\mathbf{A}$ may not be diagonalizable, and we consider generalized eigenvectors. Specifically, we index (generalized) eigenvectors such that each eigenvector $\mathbf{v}_i$ with generalized eigenvectors, $\mathbf{v}_{i+1}, \ldots, \mathbf{v}_{i+L_i}$ forms its generalized eigenvector chain of length L_i – i.e., for $0 \le l \le L_i$, $\mathbf{v}_{i+l}$ represents l-th element of the chain. By representing $\mathbf{A} = \mathbf{V}\mathbf{J}\mathbf{V}^{-1}$ and $\Delta \mathbf{e}_k = \mathbf{V}[\alpha_1 \ \ldots \ \alpha_n]^T$, where $\mathbf{V} = [\mathbf{v}_1 \ \ldots \ \mathbf{v}_n]$ and $\mathbf{J}$ is Jordan form of $\mathbf{A}$, we can exploit the property of Jordan block matrices [4], to obtain the following

$$\mathbf{A}^j \mathbf{v}_{i+L_i} = \sum_{l=0}^{min(j,L_i)} \binom{j}{l} \lambda_i^{j-l} \mathbf{v}_{i+L_i-l}.$$

This allows us to represent (70) as

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{j=0}^{k-1} \sum_{i=1}^{n} \sum_{l=0}^{min(j,L_i)} \phi_{k-j,i} \binom{j}{l} \lambda_i^{j-l} \mathbf{v}_{i+L_i-l}, \tag{73}$$

where L_i depends on the particular i that is being summed over. Let $L_{fol(i)}$ denote the number of followers of $\mathbf{v}_i$ inside its eigenvector chain (e.g., if v_i is an eigenvector $L_{fol(i)} = L_i$). Again, since $\mathbf{v}_i$ are linearly independent, we obtain

$$|\alpha_i| = |\sum_{j=0}^{k-1} \sum_{l=0}^{min(j,L_{fol(i)})} \phi_{k-j,i+l} \binom{j}{l} \lambda_i^{j-l}| \le$$

$$\le \sum_{j=0}^{k-1} \sum_{l=0}^{min(j,L_{fol(i)})} |\phi_{k-j,i+l}| \binom{j}{l} |\lambda_i|^{j-l}$$

$$\le \sum_{j=0}^{k-1} \sum_{l=0}^{min(j,L_{fol(i)})} \phi_{max} j^{L_{fol(i)}} |\lambda_i|^{j-l}$$

$$\le \begin{cases} (L_{fol(i)}+1)\phi_{max} \sum_{j=0}^{k-1} j^{L_{fol(i)}} |\lambda_i|^j, & |\lambda_i| \ge 1 \\ (L_{fol(i)}+1)\phi_{max} \sum_{j=0}^{k-1} j^{L_{fol(i)}} |\lambda_i|^{j-L_{fol(i)}}, & |\lambda_i| < 1 \end{cases} \tag{74}$$

where $\phi_{max} = \max\{\phi_{max,1}, \ldots, \phi_{max,n}\}$. If we use the ratio test for convergence of series [1] when $|\lambda_i| < 1$, we obtain

$$\lim_{j \to \infty} \frac{(j+1)^{L_{fol(i)}} |\lambda_i|^{j+1-L_{fol(i)}}}{j^{L_{fol(i)}} |\lambda_i|^{j-L_{fol(i)}}} =$$

$$= \lim_{j \to \infty} |\lambda_i| (\frac{j+1}{j})^{L_{fol(i)}} = |\lambda_i|.$$

Thus, since $|\lambda_i| < 1$ by assumption, the series converges, and all α_i that correspond to stable eigenvalues have to be bounded. Similarly, the ratio test can be used to show that the series is divergent when $|\lambda_i| > 1$. Divergence of series for $|\lambda_i| = 1$ can be shown by substitution. Namely, from (74), when $|\lambda_i| = 1$, it follows that

$$(L_{fol(i)}+1)\phi_{max} \sum_{j=0}^{k-1} j^{L_{fol(i)}} |\lambda_i|^j = (L_{fol(i)}+1)\phi_{max} \sum_{j=0}^{k-1} j^{L_{fol(i)}},$$

which given that $L_{fol(i)} \in \mathbb{N}_0$, implies that the series also diverges for $|\lambda_i| = 1$, and thus concludes the proof. ∎

Ilija Jovanov received the Dipl. Ing. and M.S. degrees in electrical engineering from the University of Belgrade, Serbia, in 2014 and 2015, respectively.

He is a Ph.D. student at Department of Electrical and Computer Engineering, Pratt School of Electrical Engineering, Duke University, where he focuses on attack resilient control of cyber-physical systems.

Miroslav Pajic (S'06-M'13) received the Dipl. Ing. and M.S. degrees in electrical engineering from the University of Belgrade, Serbia, in 2003 and 2007, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, in 2010 and 2012, respectively.

He is currently the Nortel Networks Assistant Professor in the Department of Electrical and Computer Engineering at Duke University. He also holds a secondary appointment in the Computer Science Department. His research interests focus on the design and analysis of cyber-physical systems, and in particular security-aware distributed/networked control systems, embedded systems, and high-confidence control systems.

Dr. Pajic received various awards including the NSF CAREER Award, ONR Young Investigator Program Award, ACM SIGBED Frank Anger Memorial Award, Joseph and Rosaline Wolf Best Dissertation Award from Penn Engineering, IBM Faculty Award, as well as six Best Paper and Runner-up Awards, such as the Best Paper Awards at the 2017 ACM SIGBED International Conference on Embedded Software (EMSOFT) and 2014 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS), and the Best Student Paper award at the 2012 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS). He is an associate editor in the ACM Transactions on Computing for Healthcare (ACM HEALTH) and a co-chair of the 2019 ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS'19).